

DOCUMENT RESUME

ED 278 708

TM 870 138

TITLE EQA: Educational Quality Assessment. Commentary.
INSTITUTION Pennsylvania State Dept. of Education, Harrisburg.
Div. of Educational Testing and Evaluation.
PUB DATE Jun 86
NOTE 89p.; For 1986 data, see TM 870 139.
AVAILABLE FROM Pennsylvania Dept. of Education, 333 Market Street,
Harrisburg, PA 17126-0333.
PUB TYPE Reports - Research/Technical (143) --
Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Concurrent Validity; *Educational Assessment;
*Educational Objectives; *Educational Quality;
Elementary Secondary Education; Evaluation Methods;
Grade 4; Grade 6; Grade 7; Grade 9; Grade 11; Public
Schools; School Districts; Scores; State Programs;
Test Construction; *Test Interpretation; *Test Items;
Test Manuals; Test Reliability; *Test Results; Test
Use; Test Validity
IDENTIFIERS *Pennsylvania Educational Quality Assessment

ABSTRACT

In response to a prevailing demand for better quality education in public schools in Pennsylvania, procedures were developed to measure the adequacy and the efficiency of the educational programs in public schools. The objectives currently assessed are in the following areas: communication skills (reading and writing), mathematics, self-esteem, analytical thinking, understanding others, citizenship, arts and humanities, science and technology, work, family living, health, and environment. This manual contains the information necessary to interpret the 1986 Educational Quality Assessment (EQA) reports and to learn about the background and development of the program, the statistical bases of the reports and recommendations for releasing the data. It includes details about all five levels assessed (grades 4, 6, 7, 9 and 11). Part I contains the history and background of EQA, goals of quality education, designing the assessment package (content and item selection, matrix sampling design, reliability and validity), prediction bands, condition variables, terms and definitions, administration procedures, and instrument rationales. Part II contains a summary of the instruments (listing the rational/measurement, the subscales and/or sample items, and the response choices/scoring) and information on interpreting the EQA school report. Part III contains policies and recommendations for releasing EQA data, comparisons of pre-1985 grade 11 EQA data to current data, and item comparisons across grade levels. The EQA Teacher Questionnaire is appended.
(JAZ)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

EQA Educational Quality Assessment



The Eastern Hemlock is the official State Tree of the Commonwealth of Pennsylvania.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

R. DAYS

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☒ This document has been reproduced as
received from the person or organization
originating it.
- ☐ Minor changes have been made to improve
reproduction quality.

- Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

Commentary

ED0278708

TM 870138

EQA Educational Quality Assessment



The Eastern Hemlock is the official State Tree of the Commonwealth of Pennsylvania.

Commentary

Division of Educational Testing and Evaluation
Bureau of Educational Planning and Testing
Pennsylvania Department of Education
1986

Commonwealth of Pennsylvania
Dick Thornburgh, Governor

Department of Education
Margaret A. Smith, Secretary

Office of Basic Education
D. Kay Wright, Commissioner

Bureau of Educational Planning and Testing
Donald L. Clark, Director

Division of Educational Testing and Evaluation
J. Robert Coldiron, Chief

June 1986

Pennsylvania Department of Education
333 Market Street
Harrisburg, PA 17126-0333

The Pennsylvania Department of Education, an equal opportunity employer, will not discriminate in employment, educational programs or activities, based on race, color, religion, national origin, sex, age, ancestry, physical handicap or union membership. This policy of non-discrimination extends to all other legally protected classifications. Publication of this policy in this document is in accordance with state and federal laws, including Title IX of the Education Amendments of 1972 and Sections 503 and 504 of the Rehabilitation Act of 1973. Inquiries should be directed to Susan Mitchell, Affirmative Action Officer, 503/504 Coordinator, Education Building, 333 Market Street, Harrisburg, PA 17126-0333 (717-787-1953).

CONTENTS

	Page
INTRODUCTION	iv
PART I - THE EDUCATIONAL QUALITY ASSESSMENT PROGRAM	
History and Background of Educational Quality Assessment	1
Goals of Quality Education	3
Designing an Assessment Package	4
Content and Item Selection	4
Matrix Sampling Design.....	9
Reliability and Validity	12
Prediction Bands	18
Condition Variables	21
Some EQA Terms and Definitions	27
Administration Procedures	28
Instrument Rationales	29
PART II - INTERPRETING THE EQA SCHOOL REPORT	
Summary of Instruments	41
Interpreting the EQA School Report	47
PART III - USING EQA DATA	
Releasing EQA Data	55
Comparisons of Pre-1985 Grade 11 EQA Data to Current Data	59
Item Comparisons Across Grade Levels	60
APPENDIX: EQA Teacher Questionnaire	

INTRODUCTION

"Commentary: A systematic series
of explanations or interpretations;
something that serves for explanation..."

This manual contains the information necessary not only to interpret Educational Quality Assessment (EQA) reports but also to learn about the background and development of the program, the statistical bases of the reports and recommendations for releasing the data. It includes details about all five levels assessed (grades 4, 6, 7, 9 and 11) and is, therefore, the only interpretation manual needed for the levels represented.

A companion manual, 1986 Data, contains the statistics generated from this year's EQA administration. Again, information for all five grade levels is incorporated into a single manual for convenience of district officials and others who are interested in the "total picture" of the assessment results.

PART I
THE EDUCATIONAL QUALITY ASSESSMENT PROGRAM

HISTORY AND BACKGROUND OF EDUCATIONAL QUALITY ASSESSMENT

In the early 1960's, members of the Pennsylvania General Assembly saw the need to consolidate the Commonwealth's some 2500 school districts into a more workable number. The result was Act 299 of 1963, generally known as the School District Reorganization Act. During debate on the measure, conflicting arguments about the quality of education capable of being offered by districts of varying sizes were presented. Proponents of large district size argued that more resources were available in such districts to provide the teachers and materials for better quality education. Small district advocates, however, argued that smaller units provided more homogenous, personal education which leads to quality programs. The outcome of this debate was Section 290.1 of Act 299, which required the State Board of Education to

...develop or cause to be developed an evaluation procedure designed to measure objectively the adequacy and efficiency of the educational program offered by the public schools of the Commonwealth....The evaluation procedure shall be so constructed and developed as to provide each school district with relevant comparative data to enable directors and administrators to more readily appraise the educational performance and to effectuate without delay the strengthening of the district's educational program. Tests developed...shall be used for the purpose of providing a uniform evaluation of each school district....

As the first step in carrying out the requirements of this act, the State Board of Education appointed from among its members a Committee on Quality Education. With the help of civic and professional leaders and testing specialists, the committee formulated the state's goals of education. In considering what the goals should be, the committee felt that the school played an important role in the overall development of a student, which includes not only academic but also personal and social growth. Upon recommendation of the committee, the State Board of Education adopted the Ten Goals of Quality Education in November of 1965.

In 1967, the Pennsylvania Department of Education formed the Bureau of Quality Education Assessment (now the Division of Educational Testing and Evaluation) to translate the mandates of the act and the Goals into a workable plan of assessment. During the three-year, three-phase development plan, two basic decisions were made which have shaped the program over the years. The first of these was to provide results aggregated by school rather than to give individual student data because school-based data would better provide the program evaluation information required by Act 299. The second decision was to use background information (supplied mostly by teachers and students) to provide comparisons based on conditions that vary between school communities. Grades 5 and 11 were designated as target levels for initial test development.

When the program was finally announced in 1970, more than 60 percent (over 300) of Pennsylvania's school districts requested that their schools be assessed--more than resources could accommodate. But those districts eventually participated over the next three years, and it is noteworthy that such a large number of school districts were receptive to a voluntary assessment

of this type. In November of 1973, the State Board of Education established a timetable to assess all schools with grades 5, 8 and 11 during the school years from 1973-74 to 1975-76. More than 500,000 students participated in the assessment in that three-year period. In the spring of 1976, the State Board approved plans to review the goals and the EQA instruments, and participation in EQA again became voluntary for the next two years. The EQA instruments were revised by 1978, and the State Board reaffirmed the assessment mandate by requiring all districts to participate in EQA at least once in the five-year period between 1978-79 and 1982-83.

During this five-year period, the State Board initiated a review of the Ten Goals of Quality Education. Interested parents, citizens and lay and professional groups throughout the state became involved in the review process. On March 8, 1979, the State Board adopted a revised set of goals and an accompanying preamble (see next page). The twelve new goals contained most of the intent of the original ten but were more specifically curricular-area oriented. By assessment time in the spring of 1985, new and/or revised instruments for grades 5, 8 and 11 which related to eleven of the twelve goals (Family Living is the exception) had been developed and were included in the program for the first time. A new and expanded school report format was also introduced. In addition, participation requirements had changed. A 1984 revision of Chapter 5: Curriculum Regulations of the State Board of Education recommended that districts participate in EQA on the basis of long-range planning cycles.

The next major revision in the assessment program was precipitated by the advent of the state's first mandated competency testing program, Testing for Essential Learning and Literacy Skills (TELLS) in 1984. Because TELLs testing was required at grades 3, 5 and 8, the decision was made to change the EQA grade levels to 4, 6, 7, 9 and 11 to avoid over-testing of students in grades 5 and 8. The grade 5 EQA instruments were revised for grade 4 and slightly revised for grade 6; the grade 8 instrument package was renamed "Junior High" and used at grades 7 and 9. The grade 11 package remained unchanged. This new battery of instruments was used for the first time in the spring of 1986, when EQA and TELLs testing were conducted in the same time period.

PENNSYLVANIA'S TWELVE GOALS OF QUALITY EDUCATION

The constitution of the Commonwealth of Pennsylvania states, "The General Assembly shall provide for the maintenance and support of a thorough and efficient system of public education to serve the needs of the Commonwealth." This provision mandates a quality education for each child in the Commonwealth.

The schools have the primary responsibility for the achievement of the goals of quality education as established by the State Board of Education, but they must work in close and continuous cooperation with the family, community and other appropriate social, religious and governmental institutions to insure the highest possible achievement of the goals.

To foster achievement of a quality education, the school environment should be safe, attractive and orderly, promote a willingness to work for objectives, stimulate a readiness to continue learning throughout life and encourage the fullest possible educational development of each student.

To foster achievement of a quality education, the school program should reflect the following goals.

COMMUNICATION SKILLS Quality education should help every student acquire communication skills of understanding, speaking, reading and writing.

MATHEMATICS Quality education should help every student acquire skills in mathematics.

SELF-ESTEEM Quality education should help every student develop self-understanding and a feeling of self-worth.

ANALYTICAL THINKING Quality education should help every student develop analytical thinking skills.

UNDERSTANDING OTHERS Quality education should help every student acquire knowledge of different cultures and an appreciation of the worth of all people.

CITIZENSHIP Quality education should help every student learn the history of the nation, understand its systems of government and economics and acquire the values and attitudes necessary for responsible citizenship.

ARTS AND THE HUMANITIES Quality education should help every student acquire knowledge, appreciation and skills in the arts and the humanities.

SCIENCE AND TECHNOLOGY Quality education should help every student acquire knowledge, understanding and appreciation of science and technology.

WORK Quality education should help every student acquire the knowledge, skills and attitudes necessary to become a self-supporting member of society.

FAMILY LIVING Quality education should help every student acquire the knowledge, skills and attitudes necessary for successful personal and family living.

HEALTH Quality education should help every student acquire knowledge and develop practices necessary to maintain physical and emotional well-being.

ENVIRONMENT Quality education should help every student acquire the knowledge and attitudes necessary to maintain the quality of life in a balanced environment.

DESIGNING AN ASSESSMENT PACKAGE

Each year in March about 190,000 students in about 900 Commonwealth schools complete EQA questionnaires. But do the outcomes summarized in a school's EQA report accurately reflect student progress? To answer this, questions such as the following must be answered: Who recommended the test content? What do the tests really measure and how accurately do they measure it? How are items specified and written? This section highlights the safeguards used to produce high quality instruments and how the instruments stand up to those checks.

CONTENT AND ITEM SELECTION

Review of the 1978-1984 Package

The present EQA instrument package is the fourth in the program's history. To prepare for the large-scale revisions to the previous package, EQA staff members assembled information based on their collective experience with school personnel and on occasional surveys which were follow-ups to prior assessments. The concerns and desires that became evident were incorporated into the development stages of the present instrument package. Item analyses were also conducted on each instrument used during 1978-1984 to examine items with respect to the percentage responding to each choice, the item-to-total correlations, cultural bias, etc. Items identified as statistically weak became prime candidates for removal or alteration.

Review Committees

To provide guidance in the development of the present package, various review committees were formed. The composition of the committees varied according to the area and the extent to which an instrument was being revised or developed. For example, for Arts and Humanities, Environment and Science and Technology, where three totally new tests were being prepared, the committees had greater involvement since they were responsible for recommending objectives as well as for reviewing test items. Although the tasks engaged in by particular committees differed, the general review included reacting to the EQA instrument in use from 1978 to 1984, offering suggestions on content not covered, determining objectives to be measured, evaluating prospective items and in some instances assisting in the item writing.

An example of how a test evolved to its present content structure is provided by the mathematics area. In preparation for the 1978-1984 package, the mathematics review committee was supplied with an extensive list of objectives based on the Department's Pennsylvania Retrieval of Information for Mathematics Education (PRIMES) project. Members reacted to each objective's degree of importance and indicated (1) whether the chosen objectives should be measured at the conceptual, computational or application level and (2) the grade levels at which they should be measured. Each objective was discussed, the ratings were shared and discrepancies were resolved. The 60 most important objectives were selected, and at least two items were written per objective for each grade level. These items were later submitted to reviewers representing each grade level. The items were further refined on the basis of the reviewers' comments. The revised items for all three grade levels were then submitted to yet another individual, who reviewed them for adequacy and ap-

appropriate difficulty. After further revision, the items were field-tested with students at the appropriate grade levels. Statistical analysis provided a final basis for selecting the best items for each objective. When a review committee was convened to consider mathematics objectives and items for the present package, the members first examined the objectives and statistical history of items in the 1978-1984 instrument. After developing recommendations regarding objectives, items were reviewed and targeted for retention, refinement or replacement. New or altered items were field-tested and statistically analyzed prior to their selection for the package.

Review committees were composed of subject matter specialists at the public school, intermediate unit and university levels. Some of the members were recommended by state associations such as the Pennsylvania Council of Teachers of Mathematics (PCTM) or the Keystone State Reading Association (KSRA). The committees provided an extremely valuable service in guiding EQA instrument development and revision efforts; however, responsibility for the final instrumentation resides with the EQA staff.

Content Revisions

Based on recommendations of the review committees, the content measured was revised for the 1985 package, which was still used at Grades 5, 8 and 11. When the program was changed to measure outcomes at different grade levels in 1986, the Grade 5 instruments were used for Grade 6, the Grade 8 instruments were used for Grades 7 and 9, and the Grade 11 instruments remained intact. The Grade 5 package was scaled down by a team of fourth grade teachers for use at Grade 4. Therefore, the following content comparisons apply:

Reading Comprehension: The reading tests are similar to those in the previous package (1978 - 1984) in general content structure.

Writing Skills: The writing skills tests are similar to the previous package in that they include higher level cognitive processes that differ from typical language arts tests. They tap prerequisite skills deemed necessary for coherent written expression.

Mathematics: Like their predecessors, the present mathematics tests consist of items measuring conceptual, computational and application areas.

Analytical Thinking: These measures are structurally similar to their predecessors, which carried the title "Information Usage."

Citizenship/Social Studies: These tests were formerly called "Knowledge of Law and Government." They were enlarged considerably and include more extensive coverage of history and economics than the previous versions.

Arts and Humanities: These are instruments (new in 1985) which cover students' knowledge of people, concepts, works and theories in five areas: social and critical skills, language and literature, music, visual arts and performing arts.

Science and Technology: These tests were also new in 1985. They cover students' knowledge or comprehension of science and technology concepts or their science application or analysis/synthesis skills in four areas -- life science, earth and space science, physical science and multi-disciplinary.

Environment: These knowledge measures (new in 1985) provide coverage of five major environmental study areas -- air, land, water, resources and living things -- and one general category of interrelationships.

Health Knowledge: Health knowledge is measured at Grades 4 and 6, and the instrument is structurally similar to its predecessor.

Self-Concept in School: These instruments are identical to the "Self-Esteem" measures in the previous package. Experience revealed general satisfaction with the instruments although the name was changed to more accurately describe what they measure.

Health and Safety Practices: The Grades 4 and 6 measures are new, although the Grade 6 data can be directly compared to any 1985 Grade 5 data a school may have. The Grades 7, 9 and 11 versions are somewhat enlarged and are structurally different from the previous package.

Short measures: The areas of Understanding Others, Citizenship/Societal Responsibility, Arts and Humanities Participation, Science and Technology Attitudes, Environment Attitudes and Work Opportunities and Attitudes are measured with small sets of items. School building mean scores are not produced for these areas. Instead, item data are reported with the percentages of students locally and statewide responding to each answer choice. The items comprising the Understanding Others and Citizenship/Societal Responsibility inventories were selected from the 1978-1984 package allowing direct comparison of results with prior assessments at the item level.

Item Writing

Proposed items for instruments to be revised or developed came from many people. Teams composed of staff from EQA and, for particular areas, the Bureau of Curriculum and Instruction were charged with the responsibility of working with the review committees in writing and evaluating items. Certain general guidelines were followed in writing items regardless of the area. The vocabulary and language structure used in the items had to be (1) appropriate to the grade level, (2) free of racial or sex bias in the language used or in the content of the questions, (3) free of cultural and geographic bias and (4) grammatically correct.

Field-Testing

Most of the item tryouts were accomplished through an experimental form referred to as "Form E." Since one experimental form was included in each package of 25 test booklets used in the regular spring assessment, a substantial sample of students reasonably representative of Pennsylvania was obtained. Form E item tryouts occurred during assessments from 1981 to 1984, with approximately 1200 to 1500 students responding to Form E at each grade level each year.

In the area of analytical thinking, EQA staff performed more extensive field testing in several cooperating districts. Following the administration of the field test instrument, student interviews were conducted to evaluate understandability of the items and the logic used by students to formulate their responses.

Statistical Evaluation of the Items

Following the field-test, statistical analyses were conducted with particular attention given to (1) the percentage getting an item correct for cognitive tests or the item mean for non-cognitive instruments, (2) item-total correlation and (3) item-subscale correlation. In the latter, a subscale is a score obtained by summing across a subset of items which pertain to a particular subdivision of the test. Analyses also compared the statistical results for groups based on sex, type of community and socioeconomic level. These data were taken into account when recommending that items be accepted, revised or rejected.

Final Selection of Items

A final selection of items was made by combining the information gained from reviewers' comments, student interviews and statistical analysis of field test data. A special effort was made to preserve the content integrity of each instrument so that the items comprising the final version represent the concepts, skills or attitudinal areas specified in the original design.

Because of the matrix sampling procedure, there was yet another step in the process -- allocating items to test forms. This process used item mean or percent-correct data in conjunction with content considerations to insure that the forms were as balanced as possible. See the section on Matrix Sampling Test Design for more discussion on this process.

Attitudes and Their Measurement

Attitudes, beliefs and values are abstractions. Nevertheless, they are real enough to each individual holding them. They are typically thought of as a state of readiness or a predisposition to act or react in a certain way when faced with certain situations. A person's attitudes are always present but remain dormant most of the time. They are expressed in speech or other behavior only when the object of the attitude is perceived. A person may have strong attitudes for or against astrology but actively express them only when some issue connected with astrology arises. Attitudes are often reinforced by beliefs (the cognitive component) and attract strong feelings (the emotional component) that will lead to particular behaviors (the action tendency component).

The measurement of attitudes always involves making inferences. Since the attitudes cannot be seen or measured directly, we must infer their presence from consistencies that appear in the individual's behavior. Observing individuals across time in everyday situations is probably the best way to learn how the individual thinks, feels and acts. Clearly, this method is much too cumbersome and costly when we want to investigate the intensity and direction of attitudes of a large number of people, forcing us to rely instead on paper-and-pencil measures.

The use of paper-and pencil techniques for measuring attitudes is often questioned. These questions are directed at both the measuring instrument and the students responding to them. Critics are concerned with the possibilities that items might be misunderstood, respondents may not tell the truth, similar items may not be responded to in a consistent fashion and people may respond differently at various times. In the development of the EQA attitude instruments, these and other concerns were taken into consideration.

One factor influencing answers to attitudinal items, for example, is the response set. A common response set is the tendency of respondents to tell you what they think you want to hear (make a good impression). In their developmental stages, the scales and items used previously were checked against a special 36-item instrument designed to pick up the tendency to give socially desirable answers. This scale contained such statements as:

1. I never forget to say "please" and "Thank you."
2. Sometimes I don't like to obey my parents.
3. I always finish my homework on time.
4. Sometimes I do things I have been told not to do.

Those items and scales whose scores were found to be associated with socially desirable responses were deleted or revised to minimize the relationships found. The correlation coefficients between the final total instrument scores and a 30-item version of the social desirability scale ranged from .11 through .19 for Grade 11, which indicates that the students did not usually choose the socially desirable response. Although this procedure was not used for the present test package, it would be reasonable to expect that socially desirable responding by students is associated with responses on the current attitude scales at about the same level as found previously. In fact, the matrix sampling process, which reduces the number of items to which each student must respond, probably serves to minimize this type of problem.

An attempt was also made to minimize the effect of position bias by including both positively and negatively worded items. An estimate of response position bias can be determined by counting the number of perfect scores. To get a perfect score, the student must answer in the most positive direction for each of the separate items -- a highly questionable feat. Fewer than one percent of the nearly 30,000 respondents obtained the highest possible score on the form of the test they took.

Yes, But Can They Read It?

If students cannot read the tests, the tests cannot assess the students. Although it is impossible to control the range of verbal comprehension necessary in a program assessing over 30,000 students per grade level each year, it is feasible to develop instruments that can be read and understood by the vast majority of respondents.

In the development of previous EQA packages, examiners conducting field trials were required to submit detailed reports on the understandability of directions, item sentence structures and vocabulary for each instrument. Also, students were asked to comment on the scales and identify words or items that didn't make sense to them. Changes resulting from these data increased the readability of all EQA instruments. Reading level indices were computed for the instruments and were found to be at least a grade level below the grade level being measured. Reading levels of the present instruments were kept consistent with the previous package. Since students may ask for help from examiners on all sections except reading comprehension, reading level should not present a problem.

MATRIX SAMPLING TEST DESIGN

Matrix Sampling Procedure

From 1969 to 1977, the EQA process was based on "census" testing in which all students in the grade level being tested took all the items in the test package. In 1978, matrix sampling was introduced to the EQA package and has been used since for all or most of the measures. In matrix sampling, each student takes only a portion of the total number of items in each instrument. This is accomplished by taking the complete set of items for a given area and assigning an equal number of items to separate forms. Thus in Grade 6, where there are three forms, 20 of the 60 mathematics items are assigned to each form. For the questionnaire administration, the booklets are packaged so that every third student receives a different form. In a given Grade 6 school, about one-third of the sixth grade students receive Form A, about one-third receive Form B and a final third receive Form C. Since all forms are identical in format, administration procedures can be the same as for census testing.

To further explain the meaning of "matrix" in this context, see Figures 1 to 3. This matrix is a two-dimensional array with one dimension relating to test items (columns) and the other to students (rows). Figure 1 shows a matrix having arrays consisting of N students and K items. In conventional census testing, all students would have taken all items in the test. Figure 2 shows the type of achievement data gathered by census testing for a hypothetical nine-item test. A '1' represents a correct response and a '0' denotes an incorrect response. Figure 3 shows data resulting from a matrix sampling testing procedure. There are three items for each of three forms. The figure shows that one-third of the students received items in each of the three forms. Thus all students are tested (on one form or another) and all items are administered (to a portion of the students).

	ITEMS									
	1	2	3	.	.	.	.	.	.	K
1										
2										
3										
.										
.										
.										
.										
.										
.										
.										
.										
N										

Figure 1

A matrix showing a population of N students (rows) and a universe of K items (columns)

	ITEMS									
	1	2	3	.	.	.	.	.	.	K
1	1	0	1	1	1	0	1	1	0	
2	1	1	0	0	1	0	1	1	0	
3	0	1	1	1	0	1	0	0	1	
.	1	0	0	0	1	0	1	1	1	
.	0	1	1	1	1	1	0	1	1	
.	1	0	1	0	1	0	1	0	1	
.	0	0	1	0	1	1	0	0	1	
.	1	1	0	0	1	1	1	1	0	
.	1	0	1	0	1	0	0	1	1	

Figure 2

A typical data matrix obtained through census testing, when all students take all items

	ITEMS									
	1	2	3	.	.	.	.	.	.	K
1	1	0	1							
2	1	1	0							
3	0	1	1							
.				0	1	0				
.				1	1	1				
.				0	1	0				
.							0	0	1	
.							1	1	0	
.							0	1	1	
N										

Figure 3

A typical data matrix obtained from a matrix sampling test administration. All students are involved in the testing and all items are represented in the assessment.

Rationale for Matrix Sampling

Matrix sampling has a number of advantages over census testing. Since each student takes only a portion of the items, testing time is shorter. This is beneficial for the student; fatigue effects, built up from taking many items of a similar type for a particular area and from the sheer length of a large test package, are greatly reduced. A greater time-to-item ratio removes the tendency for some tests to be speeded. Both of these effects enhance validity. Furthermore, a shorter testing time does not intrude as much on the school's regular activities. For example, at Grade 11, only about two to three hours in one day are required for testing rather than the several days required for longer census tests.

Another strength of matrix sampling is better content coverage. An expansion of the number of items means a greater range of content. For example, the EQA mathematics test, a 30-item census measure prior to 1978, was expanded to 60 items at each grade level, which permits EQA to measure a much broader range of skills.

Assignment of Items to Forms

All forms were prepared with the same format to facilitate administration. Each form has the same number of items for a particular goal area. When items are assigned to forms, the following considerations are observed:

1. Each form should have approximately the same number of positively- and negatively-worded items in the cases of noncognitive instruments. For cognitive tests, each form should have a balance of keyed responses, and these should be different from form to form.
2. Each form should be similar in content structure. For example, in the Self-Concept in School measure, each form has approximately the same number of items depicting feelings of acceptance and about the same number tapping feelings of success. In a cognitive area, each form would have about the same number of items per subscale.

Statistical Equivalence of Forms

A statistical definition of equivalence is employed in evaluating the comparability of EQA test forms. Therefore, to be statistically equivalent, the test forms should be similar in terms of mean score, variability and internal consistency. High statistical equivalence is desirable in that its presence leads to reduced error in the matrix sampling estimate of school building mean scores. The results of the statistical equivalence checks on the current instruments are presented in the EQA Data manual.

Differences between form means are evaluated by computing an effect-size index rather than employing a significance test since virtually all differences can be found to be statistically significant when samples are extremely large. The index for effect size (Cohen, 1969) is a ratio found by dividing the observed difference between means by the common standard deviation. The differences are thus converted into standard deviation units. Differences of a quarter of a standard deviation (0.25 or less) are regarded as being of minimal consequence. Past experience indicates that forms will be equal, but current data analyses will be used to determine the existence of differences.

Estimating School Means

On the EQA School Report, a score for a school is expressed as a mean or average student score. This value is an estimate of what the school mean would be if all students had taken all the items. The example below shows a matrix of data gathered from ten students, four of whom took Form A, three who received Form B and three who responded to Form C. Each form consists of three different test items. 0's and 1's refer to incorrect and correct answers respectively. Notice that an additional column for each form, labeled TOT, refers to a total score produced by adding the scores on the individual items. Student 1 achieved a score of 2 because Items 1 and 3 were correct. Student 6, who responded to Form B, has a score of 0 because none of the answers is correct. To calculate an estimated school mean, use the following steps:

1. Determine a form score for each student
2. Determine the mean student score for each form
3. Produce a weighted estimate of a school mean based on the formula

$$\hat{X} = \frac{\sum_{i=1}^F (N_i \bar{X}_i) F}{\sum_{i=1}^F N_i}$$

Where: N_i refers to the number of students responding to a given form
 $\bar{X}_i$ refers to the mean score achieved by students responding to a given form
 F refers to the number of forms
 Σ refers to the operation of summation

Calculation of Estimated School Mean Based on Multiple Matrix Sampling Procedure

		Form A				Form B				Form C			
		1	2	3	TOT	4	5	6	TOT	7	8	9	TOT
Students taking Form A	1 2 3 4	1 0 1	2										
		0 1 1	2										
		0 1 0	1										
		1 1 1	3										
Students taking Form B	5 6 7					1 1 1	3						
						0 0 0	0						
						1 0 1	2						
Students taking Form C	8 9 10									1 0 1	2		
										1 1 1	3		
										1 1 0	2		

Mean score for students taking Form A = $2 + 2 + 1 + 3 = 8/4 = 2.00$

Mean score for students taking Form B = $3 + 0 + 2 = 5/3 = 1.67$

Mean score for students taking Form C = $3 + 2 + 2 = 7/3 = 2.33$

Estimated school mean ($\hat{X}$):

$$\hat{X} = \frac{(4(2.0) + 3(1.67) + 3(2.33))}{4 + 3 + 3}$$

$$\hat{X} = \frac{(8 + 5 + 7)}{10} = \frac{(20)}{10} = 2.0$$

$$\hat{X} = 2.0$$

RELIABILITY AND VALIDITY

General Reliability Concepts

Reliability is that characteristic of a measuring instrument which deals with consistency of scores, such as within a scale itself (internal consistency) or over time (stability). Reliability coefficients are reported as two-place decimal figures ranging from .00 to 1.00. As the instrument increases in reliability, the coefficient increases in value. Generally, reliability coefficients are interpreted as the proportion of the variance in a set of scores which is caused by variation in the examinees' true scores rather than by errors of measurement. The coefficients are derived by taking into account the length of the test and the extent to which test items contribute mutually confirming, or consistent, information.

It should be noted that the most meaningful reliability estimate pertains to the manner in which test scores are used. For example, commercial standardized achievement tests are used for making decisions about individual students. For this purpose, procedures which estimate reliability based on defining true score variance in terms of student scores are appropriate. However, it is also possible to define true score variance in terms of other units such as school mean scores. Since the EQA report is designed for school level decisions, reliability estimates which take school variation into account are particularly relevant.

Another type of information which is included here and which is related to the issue of reliability is standard error. As applied to schools (the unit of decision), the statistic of concern is the standard error of the mean, which permits one to construct with a given probability the limits within which the true score of a school will fall. Finally, this section provides information regarding the percentage of omitted items typically found for each measuring instrument.

Internal Consistency

Although the internal consistency estimate of reliability is based on the student as the unit of analysis, it is presented here because of its widespread use by and familiarity to most practitioners. The conventional KR-20 and coefficient alpha reliability formulas could not be used with the present EQA test package because of the matrix sampling process in which individual students respond to only a portion of the total number of items in a particular area. Used instead was a procedure outlined by Shoemaker (1973) which estimates the average reliability of a single item within a form and then inserts this value in the well-known Spearman-Brown formula to obtain the reliability of the scores based on the full set of items comprising the measurement of a given test area. Each school then has a set of reliability coefficients calculated uniquely for the students in that building. These reliability estimates are provided in the school report. This information enables school administrators to make judgments about how reliably their group of students responded to each area. To provide administrators with comparative reliability data, the distribution of internal consistency estimates for schools comprising the state normative sample was examined and median values obtained. The median internal consistency estimate for each area assessed is presented in the EQA Data manual.

Stability of School Means

An approach to reliability which took into account the school as the unit of analysis rather than the student was the stability of group (school) means. This was done by first dividing the student data within each school into random halves and producing, through matrix sampling formulas, a mean score for each half. Correlation coefficients between the two mean scores were then calculated. These correlations can be found in the EQA Data manual. Coefficients of this type are typically lower than internal consistency estimates. Still, the magnitude of the stability coefficients indicate acceptable reliability.

Standard Error of the Mean

Since matrix sampling methodology is used in estimation of school means, an appropriate statistical technique is used to develop a standard error of the mean. The magnitude of the standard error of the mean for a particular matrix sampling design will decrease as the number of students increases. Thus, a larger standard error of the mean will occur for smaller schools. Adding and subtracting the standard error of the mean will produce a range within which the true score of a school could be expected to fall with a probability of .68. The EQA Data manual contains a table of the median values for the standard error of the mean achieved by schools.

Percentage of Omits

Although the topic of omitted items rarely occurs in discussions of reliability, it is included here because of the concern expressed by some school administrators as to whether students were responding to most of the items and how this influenced a school's score. The EQA school report contains information regarding the percentage of items omitted for each area. The median values calculated for participating schools are shown in the EQA Data manual. Overall, the percentage of items omitted is extremely small and has little impact on the reliability of scores.

A General Definition of Validity

In general, validity refers to the extent to which an instrument accomplishes the purpose for which it is used. Since there is a variety of purposes involved in testing, the evidence gathered in support of validity must be appropriate for the intended application. Test validity, like reliability, is not an invariant property of the test nor does one study provide conclusive proof of a test's validity or lack of it. No single procedure or experimental design gives a complete picture of a test's validity. Instead, each study, in its own way, helps round out a picture of what each instrument measures and, therefore, helps interpret test outcomes.

Content Validity

Most instruments in the EQA package are cognitive, measuring skills such as reading or subject matter achievement such as mathematics or science. For these areas content validity is the most appropriate form of validity information. Content validity is concerned with the extent to which the test

content is representative of the knowledge and skills the instrument was intended to measure. Evidence of content validity relies on a judgement process conducted by experts in the area measured.

Content validity is built into the EQA test development process by having subject area committees develop a content map or blueprint outlining the general structure of a test. Content maps generally consist of two dimensions. The major content categories comprise one dimension. Using mathematics as an example, the major content categories might include whole numbers, fractional numbers, decimals, geometry, measurement and number patterns. The second dimension often focuses on the psychological process required to demonstrate the acquired knowledge or skill. These categories generally represent gradations of complexity which are often simplified versions of the hierarchical approach of Bloom's Taxonomy. Three common categories include (1) recall of general facts or principles, (2) comprehension, demonstrated by accurate paraphrasing of knowledge and drawing of inferences and (3) application which involves the correct use of a rule or generalization to a new situation. In the EQA mathematics tests items are classified according to conceptual, computational and problem solving levels.

The degree of emphasis, in terms of percentage of test items allocated to each content and process category, is determined by the content committees. Items measuring objectives subsumed within these major categories are reviewed for their appropriateness to further insure a direct linkage between items and test specifications. Through this process, content validity is enhanced. The previous section, Content and Item Selection, provides further description of the development procedures related to content validity.

Criterion Related Validity

Another type of validity evidence is obtained by correlating an instrument with another measure of the same characteristic. Finding a statistically significant correlation of sufficient magnitude between the two sets of scores provides support for concurrent validity. For example, it is reasonable to expect a substantial correlation between two measures of reading comprehension or between two self-concept scales; however, the correlation between reading comprehension and self-concept would be expected to be considerably lower since two different constructs are involved.

Criterion Related Validity Information Relevant to the 1985 Package

In addition to the process of building content validity into the cognitive areas during test development, there were a number of studies conducted on the 1978-84 package which are relevant to summarize. Some of these are studies of concurrent validity between the EQA basic skills instruments and several commercial standardized tests. Although these studies were not conducted with the current instruments, it is reasonable to expect similar results. This expectation is based on the similarity of content structure and objectives comprising the previous and current EQA package. Similarly, the studies of the analytical thinking and health knowledge tests are relevant. The non-cognitive areas measured with full scale instruments have also been investigated using the criterion related approach. Criterion related validity for the self-concept scale is clearly relevant since the current measure is identical to the one used previously. Since the health and safety practices inventories underwent fairly extensive modification, the validity study is less applicable.

Reading Comprehension, Writing Skills and Mathematics

A frequently asked question was, "To what extent do the EQA basic skills tests relate to commonly used standardized achievement tests?" To answer this question, the EQA reading comprehension, writing skills and mathematics tests were administered to students in various school districts that used the Iowa or Stanford, and in one case, the California Achievement Test. The available student total and subscale scores from the commercial tests were correlated with student scores on the appropriate EQA test. Results were similar across grade levels. Correlations for reading comprehension ranged from .58 to .84 with a median correlation of .76. For writing skills correlations ranged from .38 to .83 with a median of .72. Mathematics correlations ranged from .58 to .86 with a median of .77. A tabled presentation of these results may be found in the former EQA technical manual, *Getting Inside the EQA Inventory* (Kohr, Hertzog & Seiverling, 1984).

A study conducted by Blust and Kohr (1981) examined the relationship of building level correlations between six frequently used commercial standardized tests and the EQA package in the areas of reading comprehension, writing skills (language arts) and mathematics. Building level correlations between commercial tests administered in the Spring and the EQA tests ranged from .59 to .86 (median of .82) in reading comprehension. For writing skills or language arts the correlations ranged from .64 to .87 (median of .82). For mathematics the correlations ranged from .76 to .89 with a median value of .84.

A passage dependency study was also conducted with the reading comprehension test. This study is described at the end of this section.

Analytical Thinking

A study conducted by James Masters and James Dorwart (1980) investigated the relationship between the EQA Analytical Thinking measure (then called Information Usage) and several other measures of critical thinking. Validity support was found for the EQA instruments at all three grade levels, 5, 8 and 11, by virtue of correlating highly with accepted critical thinking instruments and with teacher ratings of student critical thinking abilities. The correlation between the EQA instrument and the Cornell Critical Thinking Test was .68 and .64 at grades 8 and 11 respectively. Correlations with the critical thinking test used in Montana's statewide assessment program were also high, .63, .62 and .73 at grades 5, 8 and 11. That instrument was reviewed by several outside authorities and was critiqued by John Fremer, then of Educational Testing Service. In 1985, the current instrument was reviewed by Robert H. Ennis, who worked for the Illinois Critical Thinking Project at the University of Illinois at Champaign-Urbana and was co-developer of the Cornell Critical Thinking Test.

Health Knowledge

In a study conducted by Carol Bellew (1979) the Grade 5 EQA health knowledge test was correlated with the health knowledge section of the North Carolina State Assessment Test. Since the content structure of the two instruments was similar, a high degree of relationship would lend support to the validity of the EQA test. Analyses revealed a correlation of .76 between the total scores. Significant correlations were also found for three clusters of items

common to the two tests. These areas were sociological health ($r=.43$), safety ($r=.53$) and personal health ($r=.64$).

Self-Concept in School

A study conducted by Robert Reynolds (1978) examined the relationship of the Self-Concept inventory (formerly Self-Esteem) with teacher ratings of the Self-Concept as a Learner Scale (SCAL). Correlations of .76, .81 and .76 were obtained between the EQA measure and the SCAL at grades 5, 8 and 11, respectively. In addition, the analyses showed significant correlations between teacher ratings of self-concept and the EQA scale. The results provided support for the validity of the EQA measure of self-concept.

Health and Safety Practices

For validation of the Grades 8 and 11 EQA health and safety practices instrument, Carol Bellew (1979) asked teachers to identify students with outstandingly high or low health practices in the areas of personal health, safety, physical fitness and usage of drugs, tobacco or alcohol. The ratings were then correlated with student scores on the EQA health practices instrument. Data analyses revealed support for the instrument's validity at both grade levels. The "high" health habits criterion group scored significantly higher than the "low" health habits group in terms of the total EQA score as well as for subscale scores based on safety, physical fitness and substance usage.

Reading Passage Dependency Study

A desirable feature of reading comprehension items is that they are passage dependent, i.e., responding to an item correctly depends on the context supplied by the reading passage. For items which are passage dependent, a student must derive information or make a proper inference based on a careful reading of the passage and then correctly answer questions based on the material. If students can answer the questions correctly without reading the passage, the test hardly assesses the ability to comprehend written materials. In this case, a test could be called "passage independent" and most likely measures some characteristic such as knowledge already possessed.

The effort to gather passage dependent information for the 1978-84 EQA reading tests was similar to the approach taken by Tuinman (1973), who administered test items with passages absent and with passages present. His approach was to compare the mean test score obtained under both conditions. Generally, the more the mean score under the passage present condition exceeds the passage absent score, the greater the passage dependency. Two special forms of the EQA reading tests were prepared. In one, the items for odd numbered passages (1st, 3rd, 5th, etc.) were presented with the passages present and the items for even numbered passages (2nd, 4th, 6th, etc.) were presented with passages absent. In the second form, the items for odd numbered passages were presented with the passages absent and the items for even numbered passages were presented with the passages present.

All 48 of the EQA reading items for grades 8 and 11 were used in the passage dependency study. At grade 5, 47 of the 48 items were used. One item was eliminated because there was no discernable way of including it without cueing the correct answer to another item. Test booklets were prepared for administration by interspersing the forms so that when handed out, half the students

would randomly receive each form. The test was administered by the teachers of cooperating school districts.

Support for passage dependency was obtained at each grade level by observing that the average percentage of items correct with the passage present was about 20 percent greater than for items presented with the passage absent. The average percent correct under the passage present condition was 55 at grade 5, 56 at grade 8 and 60 at grade 11. Under the passage absent condition the percent correct was 35, 36 and 38 respectively for grades 5, 8 and 11. If a reading test is passage dependent, it is reasonable to expect a greater amount of guessing in the absence of a meaningful context as represented by the passage absent condition. In addition to lower performance, this behavior should also translate into lower internal consistency reliability estimates. This indeed was the case since the passage present condition resulted in reliability estimates ranging from .78 to .83 while the passage absent condition reliabilities ranged from .18 to .52. In summary, the data provided clear support of the passage dependency of the EQA reading comprehension tests.

REFERENCES

(For the section Designing an Assessment Package)

- Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.
- Bellew, C. (1979). "Relationships between the EQA health tests and other measures of health knowledge and habits." Unpublished report. Harrisburg, PA: Pennsylvania Department of Education.
- Blust, R. S. & Kohr, R. L. (1981). "An examination of the relationship between school scores derived from commercial achievement tests and those from statewide assessment." Paper presented at meetings of the National Council on Measurement in Education, Los Angeles.
- Kohr, R. L., Hertzog, J. F. & Seiverling, R. (1984). *Getting Inside the EQA Inventory*. Harrisburg, PA: Pennsylvania Department of Education.
- Masters, J. R. & Dorwart, J. P. (1980). "The development and validation decision-making skills test." Paper presented at meetings of the National Council on Measurement in Education, Boston.
- Reynolds, R. N. (1978). "An investigation of the validity of the Educational Quality Assessment measure of self-esteem." Unpublished report. Harrisburg, PA: Pennsylvania Department of Education.
- Shoemaker, D. M. (1973). *Principles and Procedures of Multiple Matrix Sampling*. Cambridge, MA: Ballinger.
- Tuinman, J. J. (1973). "Determining the passage dependency of comprehension questions in 5 major tests." *Reading Research Quarterly*, 9, 206-223.

PREDICTION BANDS

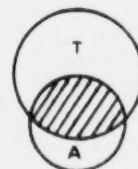
Percentiles compare building performance with all schools in the statewide normative sample but do not take into account various important conditions which affect achievement. Socio-economic level of the community, experience and educational level of the teaching staff, financial resources and type of community are a few of the variables affecting student performance. The use of prediction bands takes a number of these condition variables into account simultaneously.

How Prediction Equations Are Developed

Prediction equations are developed from a sophisticated statistical method called stepwise regression. During each step of the process each predictor variable (not all condition variables are permitted to serve as predictor variables) is statistically evaluated as to its ability to contribute to the prediction of a given score for the area measured. How much a predictor variable contributes to prediction depends on its correlation with the instrument and how highly it correlates with other variables. The addition of condition variables to the equation stops when a statistical criterion is reached, such as the ability to account for a specified amount of variance. Consequently, the number of variables comprising a prediction equation will vary from area to area. In the EQA program, prediction equations have had as few as four variables and as many as eight variables. The diagrams and explanation below illustrate the stepwise process.

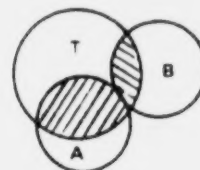
The process begins at Step 1 by entering into the equation the predictor variable having the highest correlation with the area being predicted. The overlap or proportion of shared variance between the instrument (T) and predictor variable (A) is shown by the lined area in the diagram.

Step 1



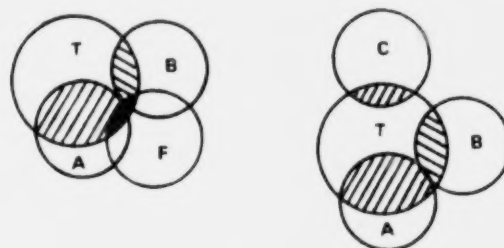
In Step 2, circles A and B both overlap with T but do not overlap with one another. Because predictor variable A and B are independent (uncorrelated) of one another they contribute the variance they share with T to the prediction of T.

Step 2



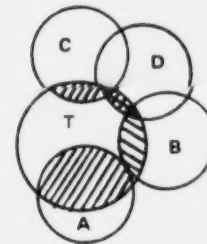
In step 3, condition variable F, which hypothetically has the next highest correlation with T, is evaluated for its ability to contribute to prediction. Variable F fails to enter the equation because all of its shared variance with T has been accounted for by A and B. Variable C does enter the equation because A, B and C are completely independent; that is, they share no variance with each other. In this instance each variable would contribute all of the variance shared with T to the prediction of T.

Step 3



In step 4, variable D is wedged between B and C; the black areas show the variance which D would contribute to T but is already accounted for by B and C. The darker shaded area represents the variance in T that is accounted for by D alone. Therefore, D enters the equation at this point.

Step 4



The development of a regression or prediction equation includes a measure of the error associated with prediction, termed the standard error of estimate. It decreases as the ability to predict increases; that is, as the correlation between the performance area and the set of variables included in the prediction equation approaches 1.0, the standard error of estimate diminishes toward zero.

How a Prediction Band Is Calculated

For each fully measured area, the EQA School Report presents a prediction band in mean scores. To demystify the process of how a prediction band is calculated uniquely for each building, consider the following illustration. Suppose that a hypothetical school building had the following results for mathematics:

Obtained school mean:	35.00
Percentile rank:	50
Stanine achieved:	5
Prediction band in mean scores:	32.45 - 36.65

Notice that the prediction band in mean score terms was 32.45 to 36.65. To calculate that band the following pieces of information must be assembled:

1. The building mean score on those condition variables comprising the prediction equation
2. Regression coefficients for condition variables comprising the prediction equation
3. A numerical value called the intercept which is part of the prediction equation
4. The standard error of estimate, also part of the prediction equation

Imagine that the prediction equation is based on PAREduc (Parental Education), PCTLI (Percentage of Low Income Students), RESIDE (Population Density of the Residential Community), MOBILITY (Frequency of Residence/School Change) and TVWATCH (Student Time Spent Watching Television). A predicted score is obtained by multiplying the value for each variable in the predictor set by the appropriate regression coefficient, summing the products and adding the intercept, as in the sample on the next page.

Predictor Variable	School Mean		Regression Coefficient		Product of School Mean and Regression Coefficient
PAREduc	3.60	x	4.20	=	15.12
PCTLI	17.00	x	-0.15	=	-2.55
RESIDE	2.00	x	2.90	=	5.80
MOBILITY	1.30	x	8.20	=	10.66
TVWATCH	1.60	x	-1.75	=	-2.80

Sum of Products 26.23

Intercept + 8.32

Predicted Score = 34.55

The prediction band is obtained by adding to and subtracting from the predicted score the standard error of estimate. In this instance suppose the estimate is 2.10:

$$\text{Lower Bound} = 34.55 - 2.10 = 32.45$$

$$\text{Upper Bound} = 34.55 + 2.10 = 36.65$$

To find the prediction band in terms of percentiles, the lower bound of 32.45 and the upper bound of 36.65 are referred to the mathematics percentile table. In this hypothetical situation the lower bound was equivalent to the 36th percentile and the upper bound reached the 75th percentile.

Cautions About Regression Equations

Regression equation information is frequently misinterpreted and misused. The following cautions should be heeded when considering regression data.

1. The entry of variables into stepwise regression does not signify that these are the only important variables. That is, variables not entering the equation may also be very important correlates of the performance area. Remember, in EQA not all condition variables are permitted to serve as predictor variables.
2. The sequence in which variables enter into stepwise regression does not necessarily reflect their relative importance.
3. The regression coefficients do not represent their significance or relative importance.
4. The variables entering into stepwise regression do not permit causal inferences nor do they lead to explanations of phenomena.
5. The process of stepwise regression is a mathematical one which does not yield a list of "similar" schools.
6. Stepwise regression is appropriate when used for purely predictive purposes.

CONDITION VARIABLES

Variable/Title	Measure	Weighting	Description
Data collected from PDE records			
1. GRENROLL Grade enrollment	The number of students participating in the assessment was read from EQA computer records.	The number of students tested in the participating grade	A higher number indicates a larger grade enrollment.
2. PCTLI Percentage of low income students	The percentage of students from low income families attending the school was obtained from PDE Chapter 1 files.	Expressed to the nearest hundredth of a percent	A higher percentage indicates that the school has a higher percentage of students from low income families.
3. TUITION Tuition rate	The tuition rate established for the school district was obtained from PDE records.	Expressed to the nearest whole dollar for the previous school year	A higher number indicates that the district reported spending more money per student.
Data collected from teacher questionnaires			
4. TSATPAR Satisfaction with relationships with parents	The teachers reported how satisfied they are with their relationships with parents and parent groups.	3 = Very satisfied 2 = Somewhat satisfied 1 = Somewhat dissatisfied 0 = Very dissatisfied	A higher score indicates that the teachers have greater satisfaction with their relationships with parents and parent groups.
5. TEDUC Teacher education	The teachers reported the level of formal education they have attained.	4 = Doctor's degree 3 = Master's degree plus 1 year 2 = Master's degree or equivalency 1 = Bachelor's degree 0 = No degree	A higher score indicates that the school's instructional staff reported higher levels of formal education.
6. TPARCONF Parent attendance at parent-teacher conferences (Grades 4 and 6)	The teachers reported the percentages of students' parents who attend scheduled parent-teacher conferences.	4 = 81-100 percent 3 = 61-80 percent 2 = 41-60 percent 1 = 21-40 percent 0 = 0-20 percent	A higher score indicates that the teachers reported a higher percentage of students' parents attend scheduled parent-teacher conferences.
7. SUPVBLDG Supervision of building (Grades 4 and 6)	The teachers indicated the position title of the person in charge of the building in which they teach and the number of buildings that person supervises.	1 = Principal of a single building 0 = All others	A score of 1 indicates that the building is supervised by a principal who is responsible for only that building; a score of 0 indicates the building is supervised by a person other than such a principal.

8. CLSIZE Average class size	The teachers reported their average class size excluding supervisory duties such as study hall.	Expressed as average class size for all teachers	A higher number indicates a larger average class size.
9. TOBSERVE Number of classroom observations	The teachers indicated the number of formal classroom observations made of their instruction each year.	4 = Four or more 3 = Three 2 = Two 1 = One 0 = None	A higher score indicates that the teachers reported having more classroom observations each year.
10. LEADER Teacher perception of building leadership	The teachers indicated the degree to which they agreed with eight positive statements about the leadership in their school.	For positively worded statements: 4 = Strongly Agree 3 = Mostly Agree 2 = Neither Agree nor Disagree 1 = Mostly Disagree 0 = Strongly Disagree	A higher score indicates that the teachers are more satisfied with the leadership in their school building.
11. TCHRNIT Teacher-initiated environment	The teachers indicated the degree to which they agreed with nine positive statements about their initiative in and control of school environment factors.	or 4 = Always 3 = Almost Always 2 = Frequently 1 = Occasionally 0 = Rarely or Never	A higher score indicates that the teachers feel they have more control over positive aspects of the school atmosphere.
12. DISRUPTN Freedom from disruptions to instruction	The teachers indicated the degree to which they agreed with two positive and six negative statements about disruptions to classroom instruction.	For negatively worded statements, the scoring is reversed.	A higher score indicates that the teachers reported fewer disruptions to classroom instruction.
13. DISCIPLN Teacher perception of discipline	The teachers indicated the degree to which they agreed with six positive and two negative statements about their perception of discipline in the school.		A higher score indicates that the teachers perceive that discipline is handled better in the school.
14. PLANNING Teacher involvement in planning functions	The teachers indicated the degree to which they agreed with seven positive statements about their involvement in various types of planning activities for the school.		A higher score indicates that the teachers feel that they are more highly involved in planning activities which take place in the school.
15. SCHLCLIM Teacher perception of school climate	The teachers indicated the degree to which they agreed with one negative and three positive statements about the general environment or climate of the school.		A higher score indicates that the teachers feel that the school has a better working environment.

Data collected from student questionnaires

16. PCTGIRLS Percentage of girls	The students indicated either male (or boy) or female (or girl).	Expressed as a percentage	A higher percentage indicates that the school has a greater proportion of girls in the participating grade.
17. PAREduc Parental education	The higher level of the following was used: (1) The students reported the highest levels of formal education attained by their fathers or male guardians. (2) The students reported the highest levels of formal education attained by their mothers or female guardians.	4 = Advanced college degree 3 = College graduate 2 = Some college, vocational, technical, business school after high school 1 = High school graduate 0 = Not a high school graduate	A higher score indicates that the school draws students from homes in which parents have higher levels of formal education.
18. RESIDE Population density of residential community	The students reported (with the help of the examiner if necessary) the types of communities in which they were living.	7 = In Philadelphia or Pittsburgh 6 = Inside a large city (100,000 to 500,000 people) 5 = Inside a medium size city (10,000 to 100,000 people) 4 = In a suburb of Philadelphia or Pittsburgh 3 = In a suburb of a large city 2 = In a suburb of a medium size city 1 = In a small town (less than 10,000 people) that is not a suburb 0 = In the open country or in a farming community	A higher score indicates that the students reside in areas of more dense population.
19. PCTWHITE Percentage of white students	The students reported which best described them: Black, White, Hispanic, Asian or American Indian	Expressed as a percentage	A higher percentage indicates that the school has a greater proportion of white students in the participating grade.
20. MOBILITY Frequency of residence/school change	The students reported the number of different school buildings they attended within the past three years because they changed residence.	4 = 5 or more school buildings 3 = 4 school buildings 2 = 3 school buildings 1 = 2 school buildings 0 = I have not moved within the past three years.	A higher score indicates that the students have changed residence and schools more often in the past three years.
21. SPARINT Student perception of parental interest in school	The students reported their opinions on two items: (1) My parents enjoy hearing about school. (2) My parents feel the school is doing a good job.	3 = Almost always 2 = Usually 1 = Sometimes 0 = Almost never	A higher score indicates that the students feel that their parents have a greater interest in school and a higher opinion of the job done by the school.

22. TVWATCH Student time spent watching television	The students reported their estimates of time usually spent watching television from the time they get home from school until they go to bed.	5 = About six hours (or more) 4 = About five hours 3 = About four hours 2 = About three hours 1 = About two hours 0 = About one hour (or less)	A higher score indicates that the students report watching more television on school nights.
23. PAREXP Student perception of parental expectations (Grades 7, 9 and 11)	The students reported their perceptions of how much schooling their parents or guardians expected them to complete.	4 = Advanced college degree 3 = Graduation from a four-year college 2 = Two-year college or other post-high school training 1 = Graduation from high school 0 = Quit high school before graduating	A higher score indicates that the students feel that their parents expect them to achieve higher educational levels.
24. EDEXPECT Student educational expectations (Grades 7, 9 and 11)	The students reported how much schooling they expect to complete.	4 = Advanced college degree 3 = Graduation from a four-year college 2 = Two-year college or other post-high school training 1 = Graduation from high school 0 = Quit high school before graduating	A higher score indicates that the students have higher educational expectations.
25. HOMEREAD Reading material in the home	The students reported the approximate numbers of magazines and books in the home.	Magazines (per month): 0 = None 1 = 1 or 2 2 = 3 or 4 3 = 5 or more Books: 0 = 0 - 24 1 = 25 - 99 2 = 100 - 249 3 = 250 or more	A higher score indicates that the students report more magazines and books in their homes.
26. TIMEREAD Time spent reading at home	The students reported how much time each day they spend reading at home.	5 = Three hours (or more) 4 = Two hours 3 = One hour 2 = 30 minutes 1 = 15 minutes 0 = None	A higher score indicates that the students report spending more time reading at home.
27. WRITEPAR Frequency of writing assignments	The students reported how often they are required to write a paragraph or more as school assignments.	4 = At least once a day 3 = At least once a week 2 = About once a month 1 = Only once or twice a year 0 = Never	A higher score indicates that the students report being required to write at least a paragraph more often.
28. PLANSWRK Perceived ability to complete schoolwork (Grade 4)	The students reported their perceptions of their ability to plan and carry out school work.	4 = Very good 3 = Good 2 = Satisfactory 1 = Fair 0 = Poor	A higher score indicates that the students perceive they have greater ability to plan and carry out schoolwork.
29. STUDYHAB Perceived quality of study habits	The students reported their perceptions of the quality of their study habits.	4 = Excellent 3 = Good 2 = Satisfactory 1 = Fair 0 = Poor	A higher score indicates that the students perceive they have higher quality study habits.

30. TIMEMATH Time spent on mathematics assignments	The students reported the approximate amounts of time each day outside of math class they spend doing math assignments.	4 = Two hours (or more) 3 = One hour 2 = 30 minutes 1 = 15 minutes 0 = None	A higher score indicates that the students reported spending more time outside of class on math assignments.
31. TESTFREQ Frequency of tests or quizzes	The students reported how often they have a test or quiz in most of their classes.	4 = More than once a week 3 = Once a week 2 = Once every two weeks 1 = Once every three weeks 0 = Once a month (or less)	A higher score indicates that the students reported having more tests or quizzes in most of their classes.
32. TESTRETN Timely return of tests (Grades 7, 9 and 11)	The students reported in how many classes the teachers return tests soon after they take them.	4 = All of my classes 3 = Most of my classes 2 = Some of my classes 1 = Few of my classes 0 = None of my classes	A higher score indicates that students reported that teachers return tests soon after administering them in more of their classes.
33. CLDISCIP Student perception of classroom discipline (Grades 7, 9 and 11)	The students reported their perceptions of discipline as a problem in the classroom.	3 = Never a problem 2 = Sometimes a problem 1 = Usually a problem 0 = Almost always a problem	A higher score indicates that the students perceive their classrooms as more free of discipline problems.
34. PCTACAD Percentage of academic/college prep students (Grade 11)	The students indicated which terms best described their present high school programs: Academic or college preparatory; General; Vocational or technical; Business or commercial	Expressed as a percentage	A higher percentage indicates that the school has a greater percentage of students in an academic or college preparatory program.
35. HRSWORK Hours of employment per week (Grade 11)	The students reported how many hours a week they work to earn money.	4 = More than 20 hours 3 = 16 to 20 hours 2 = More than 8, but less than 16 hours 1 = Some, up to 8 hours 0 = None	A higher score indicates that students reported they work more hours a week to earn money.
36. MATHINST Perception of direct instruction in mathematics (Grades 7, 9 and 11)	The students taking the class reported about how much time is usually spent on lecture and classroom discussion in mathematics class.	4 = More than 30 minutes per class period 3 = 21-30 minutes per class period 2 = 10-20 minutes per class period	A higher score indicates that the students reported receiving more direct mathematics instruction through lecture and/or classroom discussion.
37. ENGLINST Perception of direct instruction in English (Grades 7, 9 and 11)	The students taking the class reported about how much time is usually spent on lecture and classroom discussion in English class.	1 = Less than 10 minutes per class period	A higher score indicates that the students reported receiving more direct English instruction through lecture and/or classroom discussion.

38. SCIINST Perception of direct instruction in science (Grades 7, 9 and 11)	The students taking the class reported about how much time is usually spent on lecture and classroom discussion in science class.	4 = More than 30 minutes per class period 3 = 21-30 minutes per class period 2 = 10-20 minutes per class period	A higher score indicates that the students reported receiving more direct science instruction through lecture and/or classroom discussion.
39. SOCINST Perception of direct instruction in social studies (Grades 7, 9 and 11)	The students taking the class reported about how much time is usually spent on lecture and classroom discussion in social studies class.	1 = Less than 10 minutes per class period	A higher score indicates that the students reported receiving more direct social studies instruction through lecture and/or classroom discussion.
40. PCTMATH Percentage of students taking mathematics (Grade 11)	The percentage of students reporting mathematics class activity.	Expressed as a percentage.	A higher percentage indicates that more students reported that they have mathematics class.
42. PCTSCI Percentage of students taking science (Grade 11)	The percentage of students reporting science class activity.	Expressed as a percentage.	A higher percentage indicates that more students reported that they have science class.
44. INTSCHL Interest in school	The students reported their agreement with questions or statements about their interest in and satisfaction with their school situation. Grades 4,6: 12 positive questions beginning with "How do you feel...?" Grades 7,9: 19 positively-worded and 9 negatively-worded statements Grade 11: 22 positively-worded and 6 negatively-worded statements	Grades 4 and 6: 3 = Very happy 2 = A little happy 1 = A little unhappy 0 = Very unhappy For positively worded statements at Grades 7, 9 and 11: 3 = Strongly agree 2 = Mostly agree 1 = Mostly disagree 0 = Strongly disagree For negatively worded statements, the scoring is reversed.	A higher score indicates that the students are more interested in and satisfied with their school situation.
** PARENC Parental encouragement ** TEACHEX Student perception of teacher expectations	The students reported their perceptions of (1) the degree parents or guardians encourage them and (2) what their teachers expect of them.	-To be one of the best students -To be an above-average student -To be at least an average student -To do just well enough to get by (Parents) -To be a below-average student (Teachers) -I don't know	A weight was not placed on the responses, but it is generally considered that more parental encouragement and higher teacher expectations are positive attributes. "I don't know" may be the least favorable condition.
** TESTCOND Conditions at the time of testing	The students reported whether the examiner seemed positive and whether they and other students took the tests seriously.	-Strongly Agree -Mostly Agree -Mostly Disagree -Strongly Disagree	No weights were assigned to the responses. Reports of positive attitudes aid in validation of other results.

**The results for these variables are printed at the end of the school report.

SOME EQA TERMS AND DEFINITIONS

1. Matrix sampling - Matrix sampling is a testing procedure in which each student takes only a portion of the total number of items for every measured area. In EQA, this means that the items for each full measure at Grades 4 and 6 have been placed on three forms, and those for each full measure at Grades 7, 9 and 11 have been placed on four forms. This process permits broadening of the scope of the instruments while simultaneously decreasing the time needed for administration. Although the process for determining a school mean score is somewhat complicated (see the section on Matrix Sampling Test Design), the general idea is to calculate a mean score for each form and then sum the mean scores.
2. Census testing - In census testing, each student responds to all items for every measured area. In EQA, six short goal area measures, the condition variables and some miscellaneous items are census tested and item data only, not mean scores, are presented for these measures.
3. Cognitive Measure - Cognitive measures are those which measure knowledge of various areas at any level of cognitive behavior -- knowledge, comprehension, application or analysis/synthesis,
4. Noncognitive Measure - Noncognitive measures in EQA are those which collect information on students' attitudes, opinions or feelings or on their opportunities for and participation in various school courses or activities.
5. Assessment - An assessment is an appraisal of something, thus EQA is an appraisal of educational quality. The term can also be used in reference to the process, as "The assessment of 150,000 students in the spring ..."
6. EQA package - This refers to all the data collection devices connected with the assessment: the student questionnaires consisting of tests and inventories and the teacher questionnaires. One EQA package was in use from 1978 to 1984; a new package began to be used in 1985.
7. Instrument - Any testing device (see "Inventory" and "Test" below) used in the assessment. There are seventeen instruments in the EQA package.
8. Inventory - Inventories are instruments which assess attitudes or opinions or collect frequency data. There are eight inventories in the EQA package.
9. Measure - A measure is a base of comparison, so EQA has seventeen instruments which are measures of the quality of education in schools.
10. Questionnaire - The EQA package contains two types of questionnaires: the booklets containing the student assessment items and the one-page teacher questionnaire.
11. Standard score - A standard score is one that expresses test performance in terms of standard deviation units from the mean. The standard deviation is a measure of the spread or variability of scores.
12. Test - Tests are instruments which measure knowledge of certain areas. There are nine tests in the EQA package at Grades 4 and 6 and eight tests at Grades 7, 9 and 11.

13. Other Terms Defined in this Manual - Other important EQA terms are defined in special sections of this manual. Some of them are *condition variables*, *percentile*, *prediction band*, *reliability*, *stanine* and *validity*.
-

ADMINISTRATION PROCEDURES

The superintendent of each participating district appointed a representative to coordinate EQA activities at each of the grade levels. These representatives were scheduled to attend a regional workshop where their duties and responsibilities were explained. At these meetings, the EQA Representative's Handbook and the EQA Examiner's Handbooks, designed to standardize administration procedures, were discussed. In addition, the representatives were briefed on the background of EQA and the format of the results the schools would receive.

Questionnaires were administered to teachers to gain information about the school from the teachers' perspective as well as some information about the teachers themselves. Teachers responded anonymously, and all classroom teachers in a school (those spending more than 50 percent of their school time in classroom instruction) were asked to complete the questionnaires.

The parents of the students to be included in the assessment were informed of the questionnaire administration and of the purposes of the program, and they were asked to provide their children with information on their educational backgrounds to be recorded on the student answer sheets. Parents who did not wish to have their children participate could have them excluded by notifying the school principal in writing. Questionnaires were available at the school for review by interested parties.

It was recommended in the representative's and examiner's handbooks that the questionnaires be administered in small group sessions using the classroom as the test site and the teacher as the examiner. Administration took about two to three hours of student time. The examiners were provided with printed directions to standardize administration, and they were oriented to administration procedures by the representatives who had attended the regional meetings. These procedures were suggested in order to help insure optimum testing conditions and accuracy of results. As a school is forced by uncontrollable circumstances to deviate from such optimum conditions, the likelihood of inaccuracies in the results is increased.

The students wrote their names on the questionnaire booklets but not on the machine-scorable answer sheets. Students were informed that they could skip any item which they felt was too personal but that any item they skipped on a knowledge measure would be scored as incorrect. The sections of the questionnaire had no time limits for completion. The EQA Examiner's Handbook listed time ranges that might be expected for each section but with the direction that students be given ample time to finish each section.

Envelopes containing the completed teacher questionnaires and the completed student answer sheets were mailed by the representative to a private contractor for scanning, scoring, compilation and analysis.

INSTRUMENT RATIONALES

Rationales and descriptions for each of the instruments in the EQA package are presented in detail in this section. For shorter descriptions and scoring explanations, see the Summary of Instruments section which follows.

COMMUNICATION SKILLS: READING COMPREHENSION

Rationale

The goal of all reading is the comprehension of meaning. Comprehension involves the integration of a complex set of behaviors which begins with the association of a sound with a given symbol and moves through word, sentence and paragraph meaning to higher level passage-interpretation skills. Students should master such fundamental skills as word attack, phonetic analysis and structural analysis at their level of sophistication, but a general reading assessment should focus on the total reading process by means of some measure of reading comprehension.

Instrument Description

Measuring literal and inferential comprehension tasks indirectly assesses the fundamental reading skills and is perhaps the best global measure of reading achievement in a survey form with limited space. Comprehension tasks at various difficulty levels can be delineated, and the 48 items in the instruments at each grade level measure some of them:

<u>Level</u>	<u>Comprehension Tasks</u>
Literal: Comprehension of details, relationships and major ideas which are explicitly stated in a passage.	Identifying main idea Identifying supporting details Identifying cause and effect
Inferential: Comprehension of details, relationships and major ideas which are implied but not explicitly stated in a passage.	Inferring main idea Using supporting details Inferring cause and effect Predicting outcomes Drawing conclusions Determining author's point of view Making judgements Determining the (appropriate) meaning of a (multiple-meaning) word

The passages on which the items are based are either narrative or expository in nature. For example, excerpts from poetry, literature, news and weather reports and a warranty serve as passage material. Each item was reviewed to insure that it was passage dependent (could not be answered unless the student had read the passage) and that it was not biased against a cultural or racial group or against either sex.

COMMUNICATION SKILLS: WRITING SKILLS

Rationale

Communication skills include the ability to express ideas through writing. The climate of today's classroom reflects a growing demand for teachers of writing to make their instruction more effective. In addition, all teachers must emphasize writing as an integral part of the learning experiences they provide for students.

Instrument Description

There are two appropriate procedures with which writing skills can be measured: (1) an essay test which requires the student to write several short essays, letters and reports or (2) an objective test which requires the student to choose the best way of expressing something. While the essay test engages the student in actual writing, it is very time consuming and costly to administer and score.

The objective tests for writing skills used in this program measure some of the skills involved in writing standard edited English. The skills include punctuation, capitalization and usage. Also included are standard forms of organization and development of sentences and paragraphs and elements of style, tone and flavor. With carefully designed items assessing this broad range of skills, this type of measurement is highly correlated with tests of actual writing. Research has shown that students who do well on such a test tend also to be competent writers.

For Grades 7 and 9, 64 items measure four skill categories. At Grades 6 and 11, there are 60 items per instrument and Grade 4 has 48 items. Four categories of skills are assessed:

1. Mechanics and usage - recognizing the application of conventions
2. Sentence sense - recognizing clear, concise and effective sentences
3. Paragraph sense - recognizing coherence, unity and transition in the organization and development of paragraphs and larger units
4. Style, tone and flavor - recognizing effective and appropriate language for a given purpose, audience or situation

MATHEMATICS

Rationale

Mathematics instruction should include the study of a wide range of basic concepts. Students should develop the ability to handle mathematical operations or computation. In addition, development of the ability to reason logically and to respect evidence aid students in future problem solving.

Instrument Description

The mathematics tests are designed to measure the skills and concepts that are generally considered to be important for all students. The set of skills and concepts were identified using the *PRIMES* Content Authority List*. Items measure a concept itself, computation related to the concepts or problem solving involving a concept.

At each grade level, the mathematics test is a 60-item test. Items include questions about whole numbers, fractions, decimals, geometry, measurement, number patterns and other topics. There are no higher level concepts such as those encountered in trigonometry, calculus or even high school algebra. Measurement and estimation using both the English system and the metric system are included. Items are categorized as in the following examples from the Grade 11 instrument:

<u>Subscale</u>	<u>Problem Type</u>	<u>Concept/Problem</u>
Whole numbers	Problem-solving	Dividing by 10 or greater whole numbers; finding miles per gallon
Fractions	Computation	Sample: $4/5 + 7/8 =$
Decimals	Computation	Sample: $.28 \times 11.6 =$
Geometry	Problem-solving	Choosing degrees of angles of an acute angle
Measurement	Concept knowledge	Given centimetres, estimating a perimeter in metres
Number patterns	Concept knowledge	Choosing multiples of a number

*The Pennsylvania Retrieval of Information for Mathematics Educational System

ANALYTICAL THINKING

Rationale

As described in Bloom's *Taxonomy of Educational Objectives**, analysis involves breaking down material into its constituent parts and examining their relationships. This thinking ability is necessary in effectively comprehending information, in using logical processes to solve problems or make decisions and in determining the relative virtues of objects or ideas. The ability to think analytically is of the utmost importance in a society as complex as ours. Students must be taught to critically analyze information in order to make responsible decisions as citizens and in order not to be overly influenced by media appeals. When faced with a problem solving or decision making situation, they must be able to assess this situation realistically. A major skill needed is the ability to draw valid inferences from the available facts. A variety of skills are required to move from the problem definition stage to the problem solution stage. These include determining which additional information is needed before a problem solution can be found, predicting the consequences of different courses of action, using analogical thinking to draw comparisons with previously acquired knowledge, adding to current knowledge through inductive learning and evaluating the acceptability of each possible solution. When all such processing activities have been completed, students must be able to take all available information into consideration when drawing conclusions about which decision to make or problem solution to choose.

Instrument Description

These instruments are made up of short, written descriptions of problem situations involving a central character(s). Students are asked to think about the information presented and to answer multiple-choice questions about it. Information is presented in a variety of ways. These include paragraphs, tables, graphs and charts. The Grades 4 and 6 instruments have 33 items; 32 items are employed at Grades 7, 9 and 11. The instruments are composed of the following three item groups:

Making inferences items measure the ability to distinguish valid from invalid inferences drawn from the information presented.

Information processing items measure those skills needed in utilizing and adding to the available information. These include determining which additional information would be most useful to gather, gaining knowledge through inductive processes, using analogical skills to make comparisons with what was previously known, predicting the consequences of various courses of action and comparing possible decision choices with the decision maker's value system.

Drawing conclusions items measure the ability to take all available information into consideration when drawing conclusions about which decision choice to make or problem solution to choose.

*Bloom, B. S. (Ed.) *Taxonomy of Educational Objectives. Handbook I: Cognitive Domain*. New York: David McKay Company, Inc., 1956.

CITIZENSHIP/SOCIAL STUDIES

Rationale

Educating children to live in a democracy has long been an ideal for American schools. It has been a special focus of the social studies programs in particular and, most recently, the cognitive aspects of living democratically have been stressed. These cognitive areas of concern are seen as centered about knowledge of our form of government, its history (including the history of the Constitution and the economic history of the United States), the skills, rights and responsibilities inherent to effective living in a democratic environment and the economic principles which undergird our way of life.

It is seen as imperative that students be equipped with certain basic understandings if they are to become fully functioning citizens, alert to the part they can and must play in the democracy in which they live. These understandings, if mastered, are those which would enable them to meet the challenges posed by modern society.

Instrument Description

Several areas were identified as essential to effective citizenship. They are broadly described as Government, History and Economics. Government concepts deal with the practical questions of how government works and how one may have an impact on the processes of government. The tests include items dealing with (1) the need, role, function and processes of government and law, (2) leadership, power and their limitations, (3) political parties, (4) participation in citizens' rights and responsibilities and (5) political interdependence and conflict resolution.

History is broadly seen as covering (1) the development of political and economic institutions historically and (2) development of the Constitution and its amendments. In understanding the history of these bulwarks of American society, one is led to a deeper understanding of what may be changed or shaped by current processes without distorting the foundations on which these institutions rest.

In our time, economics has come to be seen as a fundamental process of every society, but one which has particular ramifications in a democracy. Knowledge of economic practices, both personal and public, is therefore required for competent functioning in our society. Economics includes (1) differing economic systems, (2) relationships of economic and political power, (3) economic interest groups, (4) economic rights and responsibilities, (5) economic interdependence and (6) basic principles of economics.

At Grades 4 and 6, students respond to 21 and 48 items respectively, demonstrating their knowledge of the above areas. At Grades 7, 9 and 11, 60 items are used with more emphasis on economics than for the lower grade levels.

ARTS AND THE HUMANITIES

Rationale

It is generally agreed that for individuals to grow to their full potential and learn to give meaning to their lives, they must develop self-awareness and a capacity for self-expression. Many arts experiences lead to the development of these attributes. But beyond individual needs are also the very pressing needs of humankind--those of creating a more equitable, just and humane world. Exposure to the humanities (the branches of learning that deal with human thought and cultures) can help to place into society individuals who are sensitive to these needs and who will perhaps act on that basis as they go about their lives in the future.

For these reasons, a quality education program should consist of a balanced curriculum addressing a wide range of educational goals which include courses in the arts and the humanities. The National Endowment for the Humanities broadly defines the humanities as including but not limited to the study of language, literature, history, philosophy, archeology, sociology and the arts. Five broad areas of arts and the humanities which are most often a part of basic education were identified for the purpose of this assessment:

Social and critical skills, including philosophy, world religions and the general concept of the humanities.

Language and literature, including works and authors of prose and poetry and the linguistics of the English language.

Music, including musical notation, musical works, composers, performers and musical instruments.

Visual arts, including art works, artists, arts processes and materials; architecture.

Performing arts, including dance, theater, plays, playwrights and performers in all areas.

State Board regulations cite five major learning objectives for this area but only two, comprehending principles and concepts and participating in aesthetic and creative experiences, can be measured within the procedural limitations of this assessment.

Instrument Description - Cognitive

Thirty multiple-choice items at Grade 4, 36 at Grade 6 and 48 items at Grades 7, 9 and 11 measure students' knowledge of people, concepts, works and theories in the five areas of the arts and the humanities listed above.

Instrument Description - Noncognitive

As a measure of opportunity and participation in arts and the humanities, a list of courses and activities is presented to the students. They are asked to indicate for each activity how often they participated in it in school in the past three years. The response choices are (1) Have not taken or participated, (2) Less than one year, (3) One year, (4) Two years and (5) Three years.

SCIENCE AND TECHNOLOGY

Rationale

The development of scientific literacy through the public schools is basic for living. Although there are many to structure science curricula, the traditional course offerings in schools are life science, earth and space science and physical science. Concepts from all three of these areas are taught at all levels.

Instrument Description - Cognitive

The domain of the science and technology tests consists of four major areas:

- Life Science
- Earth and Space Science
- Physical Science
- Multi-Disciplinary/Scientific Notions. (This area was added to permit items that contained information not unique to a single discipline.)

Items are at four levels of cognitive behavior:

Knowledge/recall, where the student identifies a correct answer from memory.

Comprehension, where the student must demonstrate an understanding of the concept.

Application, where the student must demonstrate the ability to apply scientific knowledge.

Analysis/synthesis, where the student must analyze or synthesize scientific data.

Thirty items measure these topics at Grade 4. For Grade 6, life science, earth and space science and physical science are approximately equally represented and the levels of cognitive behavior are primarily knowledge and comprehension within 36 items. Grades 7 and 9 students are given 40 items with some emphasis on physical science. Most of the items are at the comprehension and application levels. Grade 11 students respond to 44 items with a greater emphasis upon physical science than upon the other two areas. Most of the items are at the comprehension and application levels, some are at the analysis/synthesis level and few, if any, are at the knowledge/recall level. Knowledge and skills taught exclusively in high school chemistry and physics classes are not included on the test.

Instrument Description - Noncognitive

The noncognitive science and technology items consider two major components of the students' relationship to science -- the classroom and the real world. For the former, students are asked if their science classes are interesting and if the knowledge gained relates to the real world. The latter component is addressed by asking students about their perceptions of the past, present and future contribution of science to the quality of life. The six items at Grades 4, 6, 7 and 9 and the eight items in Grade 11 are printed in the appropriate school report along with the percent of students both locally and statewide selecting each of the four responses: (1) Strongly Agree, (2) Mostly Agree, (3) Mostly Disagree and (4) Strongly Disagree.

ENVIRONMENT

Rationale

The future of global quality of life depends to a great extent on human treatment of the environment--what it has been in the past and what it is at present. Therefore, since history reveals the harm done by centuries of abuse and misuse of the environment, it is imperative that all people, young and old, practice responsible stewardship of the environment today. The development of environmental awareness must begin as early as possible in children's lives and must continue throughout life.

In addition to being knowledgeable about the earth's human and natural resources, students should understand the interrelationships among and interdependence of human and natural systems, the pressures put on the earth's finite resources by the needs and demands of an ever-increasing world population, and the consequences of ignoring the need to act in a responsible manner. Students should develop responsible environmental attitudes and values and, in turn, should be motivated to use sound individual practices and to participate in group action programs designed to improve the quality of all life.

Instrument Description - Cognitive

Five major environmental study areas and one general category of interrelationships/interdependence concepts were identified for measurement:

Air, including causes and effects of pollution and how to reduce and control it; noise pollution; acid rain

Land, including conservation of farmland, solid waste disposal, composting; pollution - dumps, radioactive waste disposal

Water, Including conservation and treatment; thermal pollution, phosphates, controls; interrelationships - amount of water available

Resources, including recycling; energy items on conservation, nonrenewable fuels, insulation, future sources, solar; interrelationships - amounts of natural resources, food chains, food production

Living Things, including population; extinction; ecosystems; introduced species; insecticides; futures; zoning

General Interrelationships, including concepts of ecology, environmental stewardship, environmental problem-solving

Grades 4 and 6 students respond to 27 items and Grades 7, 9 and 11 students to 40 items demonstrating their knowledge of the six areas.

Instrument Description - Noncognitive

To obtain a measure of environmental attitudes and values, students at all five grade levels are asked to indicate how they feel about concepts described by six statements, positively or negatively worded, by choosing one of the responses: (1) Strongly Agree, (2) Mostly Agree, (3) Mostly Disagree or (4) Strongly Disagree.

HEALTH

Rationale

In their own interest, as well as in the interest of society at large, students should know how to take care of themselves and how to keep physically fit. They should know what the requirements are for physical and mental health and what practices harmful to health should be avoided.

To develop sound health-promoting behavior an individual must acquire a basic understanding of the physical, mental and social aspects of six main areas:

Safety and First Aid

Personal Health, Nutrition and Physical Activity

Family and Consumer Health

Tobacco, Alcohol, and Other Drugs

Community Health, Disease Prevention and Control

Human Growth and Development

In addition, the individual must display a willingness to consistently exhibit practices which are conducive to the maintenance of physical and emotional well-being in the above six areas.

Instrument Description - Health Knowledge

Since it is important that knowledge be established early enough to lead to the development of sound health practices, the knowledge level is measured at Grades 4 and 6. The instruments contain 33 and 51 items respectively which cover the six areas. Items deal with the problems and effects of smoking, alcohol consumption and drugs; proper practices relating to fire, electricity, bicycle riding and swimming; first aid; dental care; nutrition; disease control; exercise; digestion and general care of the body.

Instrument Description - Health and Safety Practices

Understanding how diseases and their prevention, dental care, nutrition, personal hygiene, physical exercise, safety and drug use relate to the structure and function of the human body is an important first step in each individual's willingness to consistently exhibit habits which are conducive to the maintenance of personal health and well-being. Therefore, assessment in this goal area attempts to get at students' willingness to display proper health behaviors in a variety of situations.

The student is asked to decide whether he or she would take a given health-related action. Each action is predefined as indicating either good or poor health practice. The contexts surrounding the choices are systematically varied. The number of item situations is 33 at Grade 4, 36 at Grade 6, 48 at Grades 7 and 9, and 56 at Grade 11.

SELF-CONCEPT IN SCHOOL

Rationale

Students, regardless of their grade level and pattern of talents, should experience a school environment that will enhance their self-concept. A constructive school climate can positively influence student self-confidence, sense of competency, feelings of acceptance, expectation of success and belief in their ability to exert an influence in life's affairs, all of which are important in relating to others and becoming a successful member of society.

Self-concept is a personal judgement of worthiness. It is a subjective experience which the individual conveys to others verbally or by other behavior. Most theories acknowledge that our self-image and feelings of worthiness are determined largely by how well we can live up to our own aspirations and meet expectations of others.

Aspirations become closely associated with personal goal setting behavior originating in our internalized system of values. Expectations are external in nature and are related to goals set collectively by society or by significant individuals in our lives. Assessment in this area is based on several components believed to be related to the development of positive self-concept in the school context.

The first is a feeling of acceptance. A feeling of self-worth and acceptance by teachers and classmates is an important ingredient to the development of a positive self-concept. Those lacking a feeling of acceptance often have difficulty in interpersonal relations and tend to believe that others have little confidence in, or low regard for, them.

Secondly, feelings of self-determination and success at school work, class recitation and teacher relationships help to mold feelings of self-confidence which are so important to adequate coping in daily life. Those who lack feelings of success and confidence are sometimes characterized as cautious, submissive individuals who feel inadequate, are fearful and expect to be unsuccessful in dealing with new situations.

Instrument Description

The self-concept items are categorized into three major groupings or clusters: (1) Teacher Relationships, (2) Peer Relationships and (3) School Self-Image. For each grade level, the self-concept instrument is comprised of 32 or 33 short, self-descriptive statements reflecting the school context. Response options available to the student are (1) Strongly Agree, (2) Mostly Agree, (3) Mostly Disagree and (4) Strongly Disagree.

UNDERSTANDING OTHERS

Rationale

Students should acquire attitudes which will facilitate easy interaction with all people regardless of skin color, religious beliefs, physical characteristics, socioeconomic status or cultural tradition. Schools, although not solely responsible for the development of these attitudes, should be aware of the degree of tolerance exercised by their students and should promote positive interpersonal relations in the school setting.

Instrument Description

Students are presented with descriptions of two school situations involving interactions with others. They are asked if they would approach or tolerate individuals who differ from themselves in five ways: economic status, religion, race, physical appearance or cultural tradition. The instrument for Grade 4 contains twelve items and those for the other grades contain ten items. The response options available to students are (1) Yes, (2) I think so, (3) I don't know, (4) I don't think so and (5) No.

CITIZENSHIP/SOCIETAL RESPONSIBILITY

Rationale

Teachers, administrators, curriculum developers and other concerned citizens should encourage students to assume responsibility for their actions and the actions of groups to which they belong. Opportunities should be provided for students to work cooperatively and demonstrate integrity when dealing with others. Students should be given the chance to take the initiative and assume leadership as well as support group efforts as followers. Actions which infringe on the rights or property of others, ignore the representative government which has established the laws or abandon one's personal development must be discouraged.

Instrument Description

For Grades 4 and 6, seven items ask the students to indicate their likely behavior in situations in which they could act in responsible or irresponsible ways, such as returning a lost wallet which they have found. For Grades 7, 9 and 11, students are given three situations and asked to indicate whether they are willing to behave in a positive fashion or refrain from negative actions given two different circumstances. Therefore, there are six items at these three grade levels. Response options available to students at all levels are (1) Yes, (2) Maybe and (3) No.

WORK

Rationale

In order to understand future career opportunities, a student must develop an awareness of personal preferences, interests and abilities as well as acquire a knowledge of labor market conditions and the duties, training and educational requirements of various occupations. Also necessary for continuing education in career development is a knowledge of occupational clusters and of publications providing career information. Schools should provide students with guidance to seek out the kind of education and training which will lead them to careers that are best suited to their talents and interests.

In the initial stages of vocational maturity, students become aware of different kinds of work and workers. This is followed by a growing understanding of the relatedness of educational and occupational opportunities. The more vocationally mature student will actively seek information, accept personal responsibility for career decision and finally base a career choice upon a realistic appraisal of interests, achievements and abilities.

Instrument Descriptions

Seventh and ninth grade students respond to a set of eight items focusing on attitudes toward work and career planning. Eleventh grade students are presented with four sets of items for a total of 27 in all. Sets of items and the grade levels measured are:

Attitude Toward Work (Grades 7, 9 & 11)

This set of eight items* measures general attitudes toward work and career planning. Response options are (1) Strongly Agree, (2) Mostly Agree, (3) Mostly Disagree and (4) Strongly Disagree.

Opportunity To Learn Work-Related Knowledge, Skills or Habits (Grade 11)

This set of nine items asks whether the student believes that his or her school has provided sufficient opportunity to learn or develop various career-related knowledge, skills or habits. Response options are also the Strongly Agree to Strongly Disagree range.

Opportunity To Study Career-Related Issues (Grade 11)

This set of six items measures the extent to which students have had the opportunity to study career-related issues during the school year. Students are asked to indicate how often they have done each activity: (1) Not at all, (2) Once or twice, (3) Several times and (4) Many times.

Student Career Planning (Grade 11)

This set of four items deals with career planning activities carried out by the student. The items ask whether the student has received guidance in selecting courses related to a career plan, whether the school's career-resource center was visited, whether a career-interest inventory was taken and whether a career decision has been made. Response choices are appropriate multiple-choice selections.

*These items were selected or modified from the *Texas Career Education Measurement Series* (Texas Education Agency, 1977).

PART II
INTERPRETING THE EQA SCHOOL REPORT

[illegible]

MATRIX SAMPLED, ALL DATA

(Cognitive) 4:48/6:48/7,9:48/11:48

Word meaning/contextual analysis: Students must use context clues to infer the (appropriate) meaning of a (multiple-meaning) word.

(Cognitive) 4:48/6:60/7,9:64/11:60

Style, tone and flavor: recognizing effective and appropriate language for a given purpose, audience or situation

(Cognitive) 4:60/6:60/7,9:60/11:60

Each item has four multiple choice response options. A score of one point is given for each correct response.

ANALYTICAL THINKING

(Cognitive) 4:33/6:33/7,9:32/11:32

The ability to critically analyze information is of utmost importance to successful functioning in a complex society. Given short descriptions of problem situations involving a central character, students respond to items based on information presented in paragraphs, tables, graphs or charts.

Making inferences: the ability to distinguish valid from invalid inferences

Information processing: those skills needed in using and adding to available information

Drawing conclusions: the ability to take all available information into consideration when making a decision or solving a problem

Each item has four multiple choice response options. A score of one point is given for each correct response.

CITIZENSHIP/SOCIAL STUDIES

(Cognitive) 4:21/6:48/7,9:60/11:60

Basic understandings of government and how it works, history and its meanings and the fundamental processes of economics are imperative to become a fully functioning citizen. Students respond to items demonstrating their knowledge of three broad areas.

Government: the need, role, function and processes of government; leadership, power and limitations; political parties; participation in citizens' rights and responsibilities; political interdependence; conflict resolution.

History: the historical development of political and economic institutions; the development of the Constitution and its amendments.

Economics: differing economic systems; relationships of economic and political power; economic interest groups; economic rights and responsibilities; economic interdependence; basic principles of economics.

Each item has four multiple choice response options. A score of one point is given for each correct response.

ARTS AND THE HUMANITIES

(Cognitive) 4:30/6:36/7,9:48/11:48

Arts experiences lead to the development of self-awareness and a capacity to self-expression, which are necessary to help individuals to grow to their full potential. Exposure to the humanities can develop awareness of and sensitivity to the needs of society. Students respond to items demonstrating their knowledge of people, concepts, works and theories in five areas.

Social and critical skills: philosophy, world religions, the general concept of the humanities

Language and literature: works and authors of prose and poetry, the linguistics of the English language

Music: musical notation, musical works, composers, performers, musical instruments

Visual arts: art works, artists, arts processes and materials, architecture

Performing arts: dance, theater, plays, playwrights, performers in all areas

Each item has four multiple choice response options. A score of one point is given for each correct response.

SCIENCE AND TECHNOLOGY

(Cognitive) 4:30/6:36/7,9:40/11:44

The development of scientific literacy through the public schools is basic for living, working and decision making in our technological society. Students respond to items demonstrating their knowledge or comprehension of science and technology concepts or their science application or analysis/synthesis skills.

Life science

Earth and space science

Physical science

Multi-disciplinary

Each item has four multiple choice response options. A score of one point is given for each correct response.

ENVIRONMENT

(Cognitive) 4:27/6:27/7,9:40/11:40

Environmental awareness is important to the development of practices which will lead to the maintenance and promotion of global quality of life. Students respond to items demonstrating their knowledge of five major environmental study areas and one general category of interrelationships.

Air: pollution causes and effects; noise pollution; acid rain

Land: conservation; waste disposal; pollution

Water: treatment; pollution; the water cycle.

Resources: recycling; energy sources such as fossil fuels and solar; insulation; food chains; food production

Living things: human populations; ecosystems; insecticides; introduced species; zoning

Interrelationships: ecology; environmental stewardship; problem solving.

Each item has four multiple choice response options. A score of one point is given for each correct response.

HEALTH KNOWLEDGE

(Cognitive) 4:33/6:51

To develop sound health-promoting behavior, an individual must acquire a basic understanding of various health areas. Students respond to items about the requirements for physical and mental health and what practices harmful to health should be avoided.

Safety and first aid

Personal health, nutrition and physical activity

Family and consumer health

Tobacco, alcohol and other drugs

Community health, disease prevention and control

Human growth and development

Each item has four multiple choice response options. A score of one point is given for each correct response.

SELF-CONCEPT IN SCHOOL

(Noncognitive) 4:33/6:33/7,9:32/11:32

A constructive school climate can influence the development of a positive self-concept which is important in relating to others and to becoming successful members of society. Students respond to short self-description statements about their feelings of acceptance by teachers and classmates and of self-determination in school work, class recitation and teacher relationships.

The items are divided into three major groupings or clusters: (1) teacher relationships, (2) peer relationships and (3) school self-image.

SAMPLE ITEMS:

I feel that my teachers understand me.

I get along well with kids in my class.

I often become discouraged in school.

Response Choice Weightings

	Positively Worded Item	Negatively Worded Item
Strongly Agree	3	0
Mostly Agree	2	1
Mostly Disagree	1	2
Strongly Disagree	0	3

Highest possible score: 4:99/6:99/7,9:96/11:96

HEALTH AND SAFETY PRACTICES

(Noncognitive) 4:33/6:36/7,9:48/11:56

In addition to acquiring a basic knowledge of sound health behaviors, the individual must display a willingness to consistently exhibit practices which are conducive to the maintenance of physical and emotional well-being. Students respond to statements demonstrating their willingness to display proper health behaviors in a variety of situations.

Subscales are the same as in Health Knowledge above.

SAMPLE ITEMS:

The school is offering free breakfast each morning. If I didn't have breakfast at home, I would eat breakfast at school.

While at home alone, I get a headache. I go to the medicine cabinet and find some headache pills. I would take one of the pills.

Response Choice Weightings

	Positive Action	Negative Action
Definitely Would	3	0
Probably Would	2	1
Probably Would Not	1	2
Definitely Would Not	0	3

Highest possible score:
4:99/6:108/7,9:144/11:168

SHORT MEASURES

CENSUS TESTED, ITEM DATA ONLY

UNDERSTANDING OTHERS

(Noncognitive) 4:12/6:10/7,9:10/11:10

Students should acquire attitudes which will facilitate easy interaction with all people. Students are presented with descriptions of school situations involving interactions with others who differ from themselves in five ways, and they are asked whether they would want to interact with each of the differing others.

The student differs from the respondent in one of five ways: economic status, religion, race, physical appearance or cultural tradition.

SAMPLE ITEM:

You and another student must do a school project together. Would you mind if the person who works with you is...

a student much richer than you?

a student who belongs to a religion which you do not belong to?

Response Choices

Yes
I Think So
I Don't Know
I Don't Think So
No

CITIZENSHIP/SOCIETAL RESPONSIBILITY

(Noncognitive) 4:7/6:7/7,9:6/11:6

Students should be encouraged to assume responsibility for their actions and the actions of groups to which they belong. Given varying social situations, students are asked to indicate whether they are willing, in such a situation, to behave in a socially responsible (positive) fashion or to refrain from negative actions.

SAMPLE ITEM:

A student is with friends in the hall. The student sees a new student standing alone. In this situation, I would have my friends meet the new student...
even if my friends would be upset with me.
if the student had a hard time making friends.

Response Choices

Yes
Maybe
No

ARTS AND HUMANITIES PARTICIPATION

(Noncognitive) 4:11/6:11/7,9:19/11:22

Experiences in the arts and in humanities activities and the development of skills through participation lead to the development of individuals who can take a more complete role in society. Students are presented with a list of courses and activities and they are asked to indicate how often they participated in each class or activity in school during the past three years.

SAMPLE ITEMS, GRADES 7 & 9:

Indicate how many years you have taken or participated in the following activities IN SCHOOL in the last three years:

Art class
Art history or appreciation class
Music class
.....
Band or band front
Orchestra
.....
Latin class
French class
.....
Film-making or TV production

Response Choices

Have not taken or participated
Less than one year
One year
Two years
Three years

SCIENCE AND TECHNOLOGY ATTITUDES

(Noncognitive) 4:6/6:6/7,9:6/11:8

Students should be made aware of the important contributions of science and technology to the quality of life. Students respond to items about their relationship to science in the classroom and the real world.

SAMPLE ITEMS:

I often feel successful about the work I do in science classes.
I feel that science will eventually solve most problems, such as pollution, disease and energy shortages.

Response Choices

Strongly Agree
Mostly Agree
Mostly Disagree
Strongly Disagree

ENVIRONMENT ATTITUDES

(Noncognitive) 4:6/6:6/7,9:6/11:6

Positive environmental attitudes are necessary to the development of responsible environmental practices. To obtain a measure of environment attitudes and values, students are asked to indicate how they feel about statements about the environment.

SAMPLE ITEMS:

No matter how much we harm nature it can always heal itself.
Past treatment of the environment affects us today.

Response Choices

Strongly Agree
Mostly Agree
Mostly Disagree
Strongly Disagree

WORK

(Noncognitive) 7,9:8/11:27

In order to understand future career opportunities a student must develop an awareness of personal preferences, interests and abilities. Students are asked, in a variety of formats, about their attitudes toward work and career planning and their career development opportunities and experiences.

SAMPLE ITEMS:

Grades 7, 9 and 11:

I feel that working will give me a high sense of accomplishment.

Grade 11:

How often have you studied how jobs might change in the next ten years?

Response Choices

Strongly Agree
Mostly Agree
Mostly Disagree
Strongly Disagree

OR

Not at all
Once or twice
Several times
Many times

OTHER MEASURES

CENSUS TESTED, ITEM DATA ONLY

COMPUTER LITERACY

(Cognitive and noncognitive) 4:15/6:16/7,9:11/11:12

Computers will continue to play a major role in the lives of the students. Students should have a basic knowledge of computers, some hands-on experience with computers and a positive attitude toward computer usage. Students are asked to respond to knowledge items, asked about the use they have been able to make of computers, and about their attitudes toward computer usage.

SAMPLE ITEMS (noncognitive):

Computers will make my life better.

I would enjoy learning more about computers.

Each cognitive item has four multiple choice response options.

Three noncognitive items have tailored response sets, and four noncognitive items (see the sample items) have the following range of response choices:

Strongly Agree
Agree
Don't know
Disagree
Strongly Disagree

INTERPRETING THE EQA SCHOOL REPORT

School Report Page 1

1. Area

This column lists the names of the goal areas measured by full scales, with the cognitive preceding the non-cognitive. Health knowledge is measured only at Grades 4 and 6. The sequence is similar to that of the goals as listed by the State Board of Education, but each district is encouraged to establish its own goal priority order.

2. Number students

The number of students who were given scores for each area are presented because these numbers may differ for the various areas. The differences may be due to the following:

- If the assessment was administered over a two-day period, some students may have been absent on one of the two days.
- EQA scoring protocol requires that a student complete a minimum number of items (about 40% in a cognitive area) in order for a score to be computed and to be used in the determination of the school score. (If a student responds to the minimum number of items or more, any items skipped are scored as incorrect. A valid "0" score would be one where a student answered the minimum or more of the items and had none correct.) For noncognitive areas, a score is not computed for a student who does not respond to 80% or more of the items. (If a student responds to 80% but not to 100% of the items, a total score is prorated by averaging responses on the items answered. Therefore, there is no penalty for skipping a few noncognitive items.)

3. Obtained School Score

This is the estimated school mean based on student scores using the matrix sampling procedure explained in the section on matrix sampling. The school mean scores are raw scores. For cognitive areas they are, in fact, the average of the number of correct responses the students achieved. For example, a school score of 35.89 on Mathematics can be interpreted as the students averaging 35.89 items correct out of 60. However, this is not comparative data since it does not take into account the difficulty of the items. For noncognitive areas, the school scores are the average of the students' scores (the sums of the item weights). The highest possible score equals the number of items times the highest weight. For example, for Grade 11 Self-Concept in School the highest possible score is 3 (the highest weight) times 32 items or 96.

GRADE WEIGHTED AVERAGES OF SCORES OF 901		
AREA	NUMBER STUDENTS	OBTAINED SCHOOL MEAN
COGNITIVE		
READING COMPREHENSION	99	28.93
WRITING SKILLS	99	35.64
MATHEMATICS	99	30.99
ANALYTICAL THINKING	99	17.33
CITIZENSHIP/SOCIAL STUDIES	99	21.45
ARTS & HUMANITIES	99	23.43
SCIENCE & TECHNOLOGY	99	29.41
ENVIRONMENT	99	20.16
HEALTH KNOWLEDGE	Not measured at this grade	
NON-COGNITIVE		
SELF-CONCEPT IN SCHOOL	98	56.90
HEALTH & SAFETY PRACTICES	99	105.49

However, it is important to remember that, when comparing a school score to that of another school, small score differences can translate into large differences in percentile rank. This is because the school score is an average of student scores and the distribution (or spread or variance) of these averages decreases as the size of the group increases. Therefore, the range of school scores is more limited than the range of student scores. Likewise, the range of district scores is more limited than the range of school scores. Figure 4 illustrates this phenomenon in a theoretical distribution. The actual distribution of the school scores and condition variables scores is presented in the EQA Data manual.

Figure 4 illustrates the principle that, particularly at the center of the distribution, a change in scores from A to B represents a large area under the peaked curve (i.e., district means) and hence translates into a large difference in percentile rankings. It represents a slightly smaller amount of range for school scores and a much smaller portion of the student score range. However, this phenomenon should not concern the interpreter, for such increases are difficult to achieve in schools of average or large size or in a district because the reliability of grouped scores increases in direct relationship to the group size.

4. Percentile Rank

In order to present comparative data, a percentile framework is established annually for the statewide sample of schools (see discussion in the EQA Data manual). Each of the achieved school means of the sample schools is placed in rank order. The score which appears at the lowest point is the first percentile; the score which appears at a point which encompasses 2% of the schools is the second percentile. This continues up the scale with the score at the midpoint being the 50th percentile, or state median, and the highest score being the 99th percentile.

Therefore, a percentile of 75 means that 75% of the schools in the state (or, more precisely, the statewide sample) scored lower on the measure than the schools at the 75th percentile. It does *not* mean that the students of the school answered 75 percent of the items correctly.

It is extremely important to note that the percentile ranks are based on the distribution of school scores. They may appear to be discrepant from percentile ranks on published standardized achievement tests. Several factors might account for the apparent discrepancy, such as variations in content assessed by the two tests, time of year of sampling or state versus national sampling. However, such variations are more likely to result from the derivation of the percentiles. While EQA scores are based on the distribution of school scores, the published tests' percentile ranks are based upon a distribution of student scores. Individual student scores should be compared with other student scores and schools should be compared with other schools. Yet test scoring firms often report what appear to be school scores in terms of either the average or median student percentile, i.e., one based on the student distribution with its wider range of scores.

AREA	STATE COMPARISON	
	FILE RANK	STATE
COGNITIVE		
READING COMPREHENSION	30	4
WRITING SKILLS	14	3
MATHEMATICS	5	2
ANALYTICAL THINKING	26	4
CITIZENSHIP/SOCIAL STUDIES	9	2
ARTS & HUMANITIES	12	3
SCIENCE & TECHNOLOGY	11	2
ENVIRONMENT	15	3

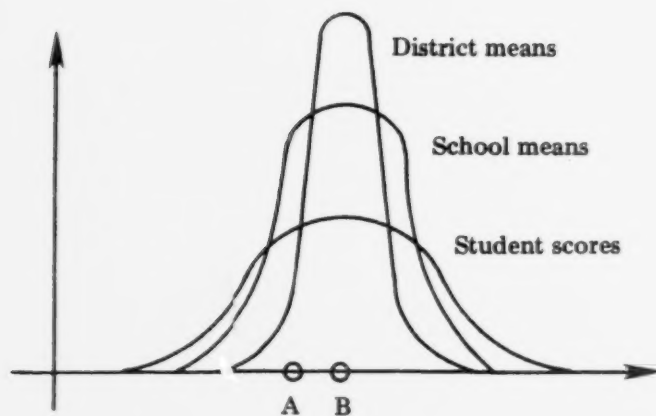


Figure 4
Score Distributions

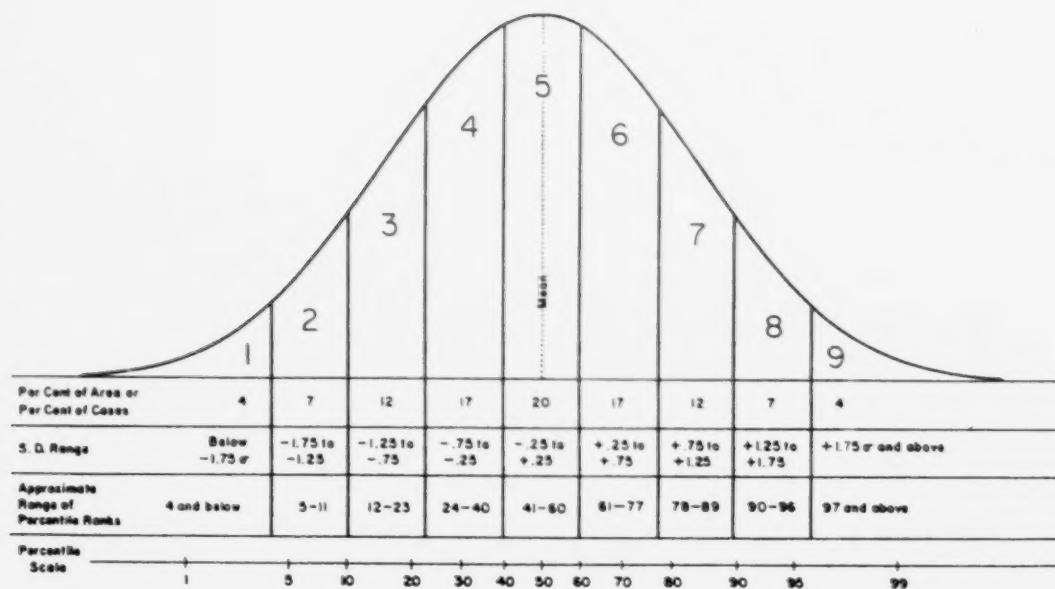


Figure 5
Stanines and the Normal Curve

5. Stanine

Stanines provide another way of comparing scores with those of other schools in the state. Stanines are like percentile ranks except that stanines are on a nine-point scale. Each school's score is placed into one of nine groups from the lowest group (stanine 1) to the highest group (stanine 9). The stanine presents a type of score that can be easily compared longitudinally. When making these comparisons, an increase or decrease of two or more stanines from one year to the next represents a significant difference. The stanine is a standardized score having a mean of 5 and a standard deviation of 2. With the exception of stanines 1 and 9, which cover the tails of a distribution, each stanine incorporates a band of scores equivalent to about one-half a standard deviation unit.

The relationship between stanines and percentile ranks is shown in Figure 5 and is as follows: Stanine 1 contains the lowest 4 percent of scores, that is, scores in percentiles 1 to 4; stanine 2 contains the next 7 percent of scores, that is, from percentiles 5 to 11; stanine 3 contains the next 12 percent of scores; stanine 4 contains the next 17 percent of scores; stanine 5 the next 20 percent; stanine 6 the next 17 percent and so on systematically through 9. From this relationship it is clear that the middle stanine of 5 contains the statewide mean score. Therefore, stanines lower than 5 indicate, on cognitive measures, a mean score less than the state average, while stanines greater than 5 indicate a mean score higher than the state average. Stanines are relatively easy to use because they are all one-digit numbers. However, the information provided by stanines is less precise than that provided by percentile ranks.

6. Obtained/Prediction Comparison

This column presents a comparison which takes into account the constraining or advantageous conditions under which schools operate. The procedure for calculating the prediction bands is explained in the Prediction Bands section of this manual. The actual school mean score is compared to the prediction band for each area measured, and the corresponding word--Above, Within or Below is printed in this column.

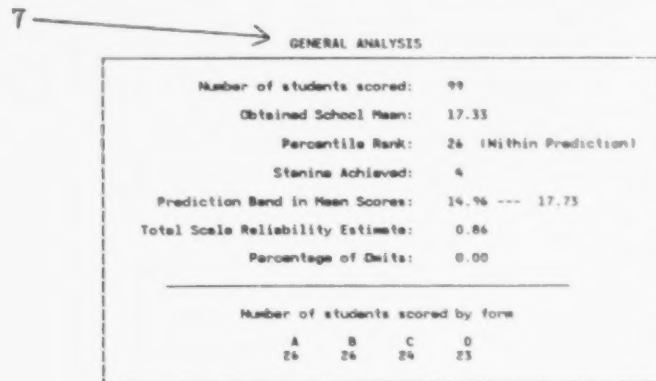
AREA	PREDICTION COMPARISON
COGNITIVE	
READING COMPREHENSION	WITHIN
WRITING SKILLS	WITHIN
MATHEMATICS	WITHIN
ANALYTICAL THINKING	WITHIN
CITIZENSHIP/SOCIAL STUDIES	WITHIN
ARTS & HUMANITIES	WITHIN
SCIENCE & TECHNOLOGY	BELOW
ENVIRONMENT	BELOW
HEALTH KNOWLEDGE	Not measured at this grade
NON-COGNITIVE	
SELF-CONCEPT IN SCHOOL	BELOW
HEALTH & SAFETY PRACTICES	WITHIN

School Report Pages 2 through 28

For each of the fully measured areas listed on Page 1, there are two pages containing all derived data. Following these, percents responding to all items in the short census measures are presented.

7, 8. General Analysis and Subscales

This data block contains the number of students validly scored, the mean (school) score, percentile rank, stanine, prediction band in mean scores, a reliability estimate for the measure, the percent of omits for the total measure and the number of valid student scores for each form.



SUBSCALES - These letter codes identify subscales on the next page.

MI - Making Inferences, Establishing Facts

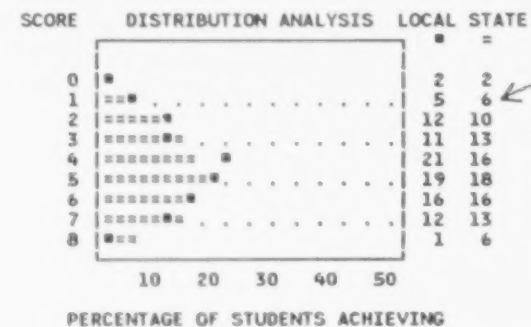
IP - Information Processing

DC - Drawing Overall Conclusions

9. Score Distributions

The second data block provides a detailed picture of how the school's students have scored; that is, how many high scores and how many low scores contributed to the school score.

The score distributions data block shows a profile of the scores for a school. The statewide distribution of student scores is also presented. The percentage of the school's students scoring at each score point (for cognitive measures) or within each score interval (for noncognitive measures) is presented as well as the percentage of students statewide.



These distributions are important since two very different groups of students could obtain the same mean school score. A school with a large number of average students and a school with about half low-scoring and half high-scoring students could have the same mean score. Certainly the future approaches to educating the students in these two schools would not be the same.

10. Item Response Data

None of the EQA instruments is comprehensive of the subject matter it represents. None will match the curriculum of a school or district perfectly. The items are only representative of the types of content that are frequently taught at or before the grade level being assessed. There is no intent for these items to drive a school's curriculum but there is valuable information to be gained by looking at the item results.

For cognitive areas, the percent of students in the school responding correctly to each item is presented with a brief description of the item. The percent of students statewide responding correctly is also given. For most of the areas, the items are labeled according to a subscale. These subscales are broad aspects of the curriculum. If the school's students fall below the statewide level on many of the items for a subscale, that broad area could be viewed as a possible weakness to be investigated in more detail. The reason(s) for the weakness may be varied. It could be an area that is not covered in the school's curriculum prior to the testing, it may not be considered important for that

school's population, or it may not have been properly taught or learned. In the latter case, methods of instruction or adherence to the school's curriculum guidelines may warrant investigation.

For the two fully measured noncognitive areas, the percentages are given for each statement with an asterisk (*) signifying the end of the continuum representing a positive approach to the subject of the item. The statewide data for each statement is also given. For the short census tested measures, item data are also presented with the asterisk at the positive response.

ITEM ANALYSIS FOR ANALYTICAL THINKING

FORM - ITEM SUBSCALE - DESCRIPTION			PERCENTAGE CORRECT	FORM - ITEM SUBSCALE - DESCRIPTION			PERCENTAGE CORRECT
A			L* S*	C			L* S*
1	MI	Drawing inferences from tabled information (data table)	42 53	1	IP	Identifying valid arguments against a conclusion	75 76
2	IP	Analyzing information to determine the consequences of a decision (data table)	62 58	2	IP	Analogical reasoning	58 51
3	DC	Reviewing all data available and deciding on an action (data table)	42 48	3	IP	Analogical reasoning	63 72
4	MI	Establishing the facts of a situation by analyzing graphic information (graph)	62 71	4	IP	Inductive reasoning	63 67
5	DC	Reviewing all data available and deciding on an action (graph)	58 68	5	IP	Inductive reasoning	50 64
6	MI	Drawing inferences from a paragraph	77 59	6	IP	Inductive reasoning	50 54
7	DC	Reviewing all data available and deciding on an action	77 78	7	IP	Analyzing information to determine the consequences of a decision (data table)	33 44
		inferences about causation		8	DC	Reviewing all data available and deciding on an action (data table)	50 49

ITEM ANALYSIS FOR SELF-CONCEPT IN SCHOOL

			Strongly Agree		Mostly Agree		Mostly Disagree		Strongly Disagree	
			L	S	L	S	L	S	L	S
(ASTERISKS INDICATE THE MOST POSITIVE RESPONSE.)										
1	B	I feel accepted by my classmates.	52	40*	40	52	4	5	4	2
2	A	My teachers expect better work from me than I am able to do.	24	13	20	26	56	49	0	12*
3	C	I am usually one of the last to be chosen for a group activity.	12	7	8	16	60	46	20	31*
4	A	My teachers listen to what I have to say.	12	23*	52	59	20	13	16	6
5	C	I feel unhappy about my schoolwork.	12	11	36	26	52	47	0	16*
6	B	I feel that my classmates do not consider me to be important.	8	7	12	18	60	47	20	29*
7	C	I am able to do many things well in school.	24	26*	64	58	8	12	4	3
8	C	I have a low opinion of my learning ability.	8	7	20	19	52	41	20	33*

11. Condition Variable Data

EQA includes prediction bands so that possibly objectionable or unfair comparisons often invited by percentiles or stanines can be tempered by the consideration of various "input" data. The input indicators used in EQA are called *condition variables* because they present information on conditions that vary from one school to another, not only in the student body but also in the teaching staff and the financial and parental support available. The charts on pages 21 to 26 of this manual list the variables collected, how they are measured and how to interpret the data presented in this part of the school report. The variable data come primarily from the students (as part of their questionnaires) and from the teachers (as responses to a separate Teacher Questionnaire). The data contain valuable information for making decisions and planning strategies for change.

The charts on pages 29 and 33 of the school report present the following information for each of the variables collected at the various grade levels:

- a. The variable name, number and acronym
- b. A raw mean score. Depending on the variable's measurement, its raw score may or may not have much meaning. For example, the score for grade enrollment (GRENROLL) is meaningful (it is the recorded number of students) but the score for school climate (SCHLCLIM) is an average of weighted scores from four statements and is very difficult to interpret. In the latter case, percentiles and item responses can present a clearer picture of the variable's meaning. Again, depending on its measurement, a high score for a variable is not necessarily good. For ex-

ample, a high grade enrollment score (GRENROLL) simply means that the school has a large number of students per grade. It does not indicate that schools having higher numbers of students per grade are better.

- c. A percentile rank. For each condition variable measured at participating grade levels, schools receive a percentile rank based on the current normative sample. This shows not only under what conditions a school is working but also its relative conditions. Like a high score, however, a high percentile does not necessarily imply "good" conditions. In addition, the norm charts in the EQA Data manual illustrate how much schools differ on a particular variable. Because these distributions use school means, the ranges of values at times can be very narrow and a small increase in a school mean value may translate into a large percentile rank change.
- d. Number replying. The number of students or teachers replying for each variable is shown as a check on the numbers of respondents on which the data are based. If numbers do not closely relate to what administrators expect, many students or teachers failed to respond to the condition variable items and the data may be skewed.

12. Condition Variable Item Results

Most of the items which were used in the calculations of variable scores are listed with their percentages of replies by students or teachers. Although the percentages for a particular school are by themselves meaningful, statewide figures are included for an additional comparison if desired.

CONDITION VARIABLE INFORMATION

VARIABLE NAME	VAR. NO.	ACRONYM	11b	11c	SCHOOL %ILE	NUMBER REPLYING
			SCORE	SCORE		
1. Grade enrollment	1	GRENROLL	99.00	26	0	
2. Percentage of low income students	2	PCTLI	41.38	95	0	
3. Tuition rate	3	TUITION	2055.00	28	0	
4. Satisfaction with relationships with parents	4 *	TSATPAR	2.23	92	30	
5. Teacher education	5	TEduc	1.50	7	30	
6. Parent attendance at parent-teacher conferences	6	TPARCONF	1.64	88	28	
7. Supervision of building	7	SUPVBLDG	- Not measured at this grade			
8. Average class size	8	CLSIZE	21.97	20	30	
9. Number of classroom observations	9 *	TOBSERVE	1.67	37	30	
10. Teacher perception of building leadership	10 *	LEADER	20.17	35	30	
11. Teacher-initiated environment	11 *	TCHRINIT	27.23	82	30	
12. Freedom from disruptions to instruction	12 *	DISRUPTN	17.03	47	30	
13. Teacher perception of discipline	13 *	DISCIPLN	20.27	86	30	
14. Teacher involvement in planning functions	14 *	PLANNING	14.10	39	30	
15. Teacher perception of school climate	15 *	SCHLCLIM	13.07	96	30	

*Items for this variable are detailed on the next pages.
 **A percentile is not appropriate for this variable.

TEACHER CONDITION VARIABLE INFORMATION

4. TSATPAR Satisfaction with relationships with parents	Satisfied				Dissatisfied			
	Very		Somewhat		Somewhat		Very	
	L	S	L	S	L	S	L	S
In your teaching situation, how satisfied are you with your relationships with parents and parent groups?	30	20	67	56	0	20	3	5

9. TOBSERVE Number of classroom observations	Percent Responding	
	L	S
None	0	3
One	40	27
Two	57	48
Three	0	10
Four or more	3	13

RELEASING EQA DATA

EQA school and district reports must be presented to the school directors and administrators (Section 290.1 of the *Pennsylvania School Code of 1949* as amended) and this should be done reasonably soon after their receipt. School results should then be presented to school staffs by their respective administrators. Suitable EQA data should be released to the community and the media. See the two recommendations sections below.

Policy for Release of EQA Data by PDE Staff

This policy will be followed by PDE staff members when a citizen or other interested person requests that the Department release a district's EQA data:

After the district has had the data for a period of time (at least three months) sufficient to provide a release to the community, any resident of the district's service area may request the release of the data from the Department. The Department will first contact the district superintendent to arrange a release to the resident. If that release cannot be provided by the district, the Department at its convenience will invite the resident and the district superintendent or designate to a presentation of the data by a member of the Division of Educational Testing and Evaluation.

Recommendations for EQA Data Release to the Media

Many districts release EQA data to the media (usually newspaper reporters) in the form of prepared news releases. This prevents reporters from misinterpreting the data and is often appreciated by them when a technical subject such as assessment information is to be presented. Two sample press releases which contain recommended topics for release are provided immediately following this section. Districts may copy these releases, substituting their data for the fictional figures, or they may otherwise use them as models for their own press releases.

Other Recommendations for EQA Data Release

1. District PR Newsletter

Often school districts will devote entire editions of their regular district newsletters to presenting EQA results. Many of these articles are longer than the press releases provided, but what is often added are charts and/or more detail on interpretation and use of the results.

2. School boards, parents and students

Unless members of the school board, parents or students are going to be involved in detailed planning groups such as those which assist districts with long-range planning, the information they receive about EQA results can consist of some explanation of the following topics. To the right of each topic is an indication of where to find information on it.

What is EQA?	HISTORY OF EQA; GOALS Commentary, pages 1 to 3
What is measured by EQA?	SUMMARY OF INSTRUMENTS Commentary, pages 41 to 46
What is in an EQA School Report?	INTERPRETING THE EQA SCHOOL REPORT Commentary, pages 47 to 53
What are our results?	GENERAL SUMMARY OF STUDENT OUTCOMES EQA School Report, page 1
Is there any other useful information in the report?	ITEM DATA EQA School Report, pages 2 to 28 CONDITION VARIABLES Commentary, pages 21 to 26 EQA School report, pages 29 to 40
What will we do with the results?	District decisions based on data analysis and discussions

3. Faculty

The faculty of each participating school should be informed by the principal or other school leader of the school's results before the parents and students but at about the same time as the school board. The topics in the above section are suitable for use with teachers as well, but more stress should be put on discussion of the details of the results and on using the data for planning or other school improvement activities. Teachers may find the item data on pages 2 to 28 of the school report especially useful.

Item data for the full measures include a subscale designation. If teachers are doing an item analysis of the Writing Skills test, for example, they may want to collect percents correct for subscale A (Mechanics and Usage) first, then for subscale B (Sentence Sense), etc. It is important, however, to view these data on two bases, criterion and normative. For criterion purposes, it is valuable to determine whether as many students answered correctly (or positively) as staff would like or expect. For normative purposes, comparisons with state percentages are possible. Remember, however, that state percentages represent the *average* and not necessarily a *high quality* school or program.

SAMPLE PRESS RELEASE 1

UPPER PENN SCHOOLS RECEIVE RESULTS FROM STATE ASSESSMENT PROGRAM

Skills in reading comprehension and mathematics are two areas in which Upper Penn School District's elementary students excel, according to the Pennsylvania Department of Education (PDE).

Reports were delivered recently by a PDE staff member who sat down with the district's administrative team to report the findings of the school assessment program conducted last spring in the Upper Penn schools as part of the state's Educational Quality Assessment program (EQA). Last March, all fourth, sixth, seventh, ninth and eleventh graders completed comprehensive questionnaires designed to measure the schools' status on the state's twelve Goals of Quality Education.

This program differs from the state-mandated TELLS program in several significant ways: it is a volunteer assessment; it measures much more than basic reading and mathematics skills; and it presents results by school rather than by individual student.

"This statewide program," says Superintendent John E. Doe, "is provided as a service to the school district. We volunteered for the program so that we could get information for future decision-making."

EQA school data are presented in a number of ways. One of them, percentile ranks, gives the district an idea about how the schools are doing compared to other schools in the state. A higher percentile means that a school scored higher than the other participating schools in the state at that level. For example, Upper Penn fourth graders scored at the 90th percentile in Reading Comprehension. This means that their score was higher than 90

percent of the other fourth grade schools in the state. Sixth, seventh, ninth and eleventh grade Reading Comprehension percentiles were the 86th, 78th, 79th, and 84th respectively, also quite good.

Other areas measured and the 4th, 6th, 7th, 9th and 11th grade percentiles were: Writing Skills: 66,70,74,81,87; Mathematics: 88,82,86,72,77; Analytical Thinking: 66,74,69,74,70; Citizenship/Social Studies: 64,60,68,78,64; Arts and Humanities: 62,61,67,87,75; Science and Technology: 43,32,56,38,43; Environment: 50,45,45, 35,59; Health: 75,78 (measured at Grades 4 and 6 only); Self-Concept in School: 80,84,71,78,77; Health and Safety Practices: 72,65,66,63,54.

The schools received other information for these areas as well as item response data (the percentages of students responding to each answer choice) for seven other areas: Understanding Others, Citizenship/Societal Responsibility, Arts and Humanities Participation, Science and Technology Attitudes, Work Opportunities and Attitudes (Grades 7, 9 and 11), Environment Attitudes and Computer Literacy.

Teachers also responded to a questionnaire concerning school conditions, and their responses are part of the school report as well. All of the information is being studied by administrators and teachers for use in decision-making at all levels of the district. "Basically, we did about as well as expected," noted Doe, "But we can always do better. I'm not happy with the scores in the science areas, so we'll be looking especially closely at those. This is tremendously valuable data."

SAMPLE PRESS RELEASE 2

UPPER PENN SCHOOLS SHOW MIXED RESULTS IN STATE ASSESSMENT PROGRAM

Upper Penn School District's board members were briefed last night on the results of last spring's Educational Quality Assessment (EQA). Superintendent John E. Doe reported that the results of the fourth, sixth, seventh, ninth and eleventh grades were "a mixed bag," and that he has already asked each school's curriculum committee to take a close look at the information and data they received.

EQA is a comprehensive assessment program in which students complete questionnaires designed to measure the schools' status on the state's twelve Goals of Quality Education. This program differs from the state-mandated TELLS program in several significant ways: it is a volunteer assessment; it measures much more than basic reading and mathematics skills; and it presents results by school rather than by individual student. The latter makes EQA data very useful for all types of planning, including district long range planning and school or district curriculum planning.

In addition, EQA provides comparisons to other schools in the state, but it also determines a prediction about how the schools should fare considering the prevailing conditions in the school community. This can prevent many unfair comparisons.

The general results reported by Doe show that in the nine cognitive, or knowledge, areas measured by EQA, the elementary schools scored well within predicted ranges in all except Arts and Humanities, where the scores were all rather low. In the junior and senior high schools, however, the Arts and Humanities scores were within predictions while the Reading Comprehension, Writing Skills and Analytical Thinking scores were all lower than was predicted.

In the senior high, Mathematics and Science scores were above predictions, which Doe claimed was a credit to the recent push to build a stronger math/science program there. He was asked if he thought that the emphasis on math and science was responsible for the lower scores in the reading and writing skills but declined to place any blame. Other areas where all schools scored within predictions were Citizenship/Social Studies, Environment, Health and Self-Concept in School.

More detailed reports will be made to the board throughout the year, Doe said, as the school committees interpret the data and compare it to the curriculum requirements in all areas.

COMPARISONS OF PRE-1985 GRADE 11 EQA DATA TO CURRENT DATA

Grade 11 is the only grade level at which comparisons to pre-1985 data can be made with confidence, and this should be done with caution, since all the instruments were revised to some extent. It is recommended that comparisons be made to *not more than the nearest 5th percentile* on the following measures:

- Reading Comprehension
- Writing Skills
- Mathematics
- Analytical Thinking (with Information Usage)
- Self-Concept in School (with Self-Esteem)
- Health and Safety Practices

In addition, item data can be compared directly in the Understanding Others and Citizenship/Societal Responsibility short measures.

However, when making even these comparisons, it is important to remember that differences may be due to a multiple of reasons other than an improvement or a decline in the educational program. Before determining that score changes are reflecting an improvement or a deficiency in program, consider these other possible reasons for changes:

1. True differences in classes of students in terms of ability levels and attitudes that influence achievement
2. Differences in composition of classes due to
 - a. Change in building structure, such as 9-12 changed to 7-12
 - b. Demographic changes such as low income or low education level families moving in or out of attendance areas
 - c. Inclusion/exclusion of special education or vocational-technical students from one assessment to another
 - d. Large changes in the percentage of girls in small enrollment schools
 - e. Large changes in school enrollment
3. Differences in test administration conditions or procedures from one assessment to another, such as
 - a. Large group administration vs. individual classroom administration
 - b. Directions given over public address system vs. by teachers when administered in individual classrooms
 - c. Presence/absence of monitors, principal or superintendent
 - d. Time of day
 - e. Local events which might influence testing climate such as pep rallies, strikes, sudden rise in local unemployment

ITEM COMPARISONS ACROSS GRADE LEVELS

To demonstrate if students are gaining certain requisite skills and knowledge as they progress through the grades, EQA assesses some content with the same items at more than one grade level. By comparing percentages of student responses across grade levels, it may be found that certain knowledge is learned by a group of students early and retained, but that the students who did not learn the information early do not learn it later. It could also be found that other knowledge is learned by more and more students as they progressed. It may even be found that a group of older students have failed to learn (or to retain) information known to a greater degree by younger students.

District and school personnel may be interested in looking at these patterns at either the district or school level. The most obvious set of comparisons can be made between Grade 7 and Grade 9 since the two packages are identical. This would be highly advisable if little difference was found between the mean scores of the two grades. Since the Grade 4 package was a refinement of the Grade 6 package, a great deal of overlap can also be found.

The items are listed by form and sequence in the instrument for the area. For example, "A 1" for Reading Comprehension is the first item on Form A of the reading test. The percent correct can be easily accessed in the EQA school report. By noting the percent correct (or responded to favorably for non-cognitive instruments) for each item, a quick comparison can be made. The statewide pattern can be obtained by accessing multiple grade level reports.

SELF-CONCEPT IN SCHOOL			INTEREST IN SCHOOL			
All items on all forms are the same at Grades 4 and 6			GRADE 4	GRADE 6	GRADES 7,9	GRADE 11
GRADE 6	GRADES 7,9	GRADE 11				
A 4		C 6	2	2 (Similar)		
C10	A 1	C 2	3	3		
	A 2	C 8	4	4		
B 9	A 3	A 7	5	5		
B 2	A 5	A 2	6	6		
	A 6	A 1	8	8		
	A 7	B 7	9	9		
C 6	A 8	C 4	10	10		
A 9	B 1	B 6	11	11		
	B 2	C 5	12	12		
	B 3	D 7			4	19
	B 4	D 6			5	8
	B 5	B 5			6	18
	B 6	A 6			7	6
	B 7	C 7			10	3
	B 8	B 3			11	4 (Similar)
A 3	C 1	B 4			12	22
	C 2	A 8			15	20
	C 3	A 4			16	24
	C 4	D 5			18	17
C 9	C 5	A 3			20	5
A11	C 6	A 5			23	25
	C 7	D 3			24	10
	C 8	D 2			27	14 (Similar)
B 1	D 1	B 1				
B 6	D 2	C 3				
C 7	D 5	B 2				
	D 6	C 1				
	D 7	D 8				
C11	D 8	B 8				

READING COMPREHENSION

GRADE 4 GRADE 6 GRADES 7,9 GRADE 11

A 1	A 1		
A 2	A 2		
	A 3	A1	
A 3	A 4		
A 4	A 5		
A 5	A 8		
A 6	A 9		
A 7	A10		
A 8	A11		
A13	A16		
	B 2	C12	
	B 3	C11	
	B 4	C 4	
B 4	B 7		
B 5	B 8		
B 6	B 9		
B 7	B10		
B 8	B11		
	B15	C 6	
	B16	C 7	
C 1	C 1		
C 2	C 2		
C 3	C 3		
C 4	C 8	D 1	
C 5	C 9	D 2	B 4
C 6	C10		
C 7	C13		
C 8	C14		
C14	B 4		
C15	B 5		
C16	B 6		
		D 3	B 6
		D 4	B 7
		D 5	D10
		D 6	D12

WRITING SKILLS

GRADE 4 GRADE 6 GRADES 7,9 GRADE 11

A 1	A 1		
A 2	A 2	B 9	A 7
A 3	A 3	C 5	
A 4	A 4	C 7	D 3
A 6	A 5		
	A 6	B 2	
	A 7	D 1	
	A 8	B 1	B 1
	A 9	C 4	D 1
	A10	B 3	
	A12	B 7	B 6
A13	A15	A10	
A14	A16		
A15	A18		
	A20	B15	
		A 2	C 3
		A 5	C 1
		A11	D 7
		A12	A10
		A13	A12
		A14	A13
		A15	B 9
B 3	B 2		
B 4	B 3	D 3	
	B 5	A 3	C 2
B 5	B 6		
B 6	C 3		
B 8	B 9		
B 9	B10	D 7	
B10	B11	A 9	B 8
B11	B12		
B12	B15	B14	
B13	B16		
B15	B18	C14	
B16	B19		
	B20	C13	
		B 5	B 4
		B 6	D 2
		B10	B10
		B11	A 9
		B12(D14)	D12
		B13	D13
		B16	D10

MATHEMATICS

GRADE 4 GRADE 6 GRADE 4 GRADE 6

A 2	A 9	B11	B 2
A 5	A 3	B12	B12
A 6	A 4	B14	B14
A 7	A 5	B15	B15
A 8	A 6	B16	B16
A 9	A 7	B17	B17
A10	A 8	B18	B18
A13	A13	B19	B19
A14	A14	C 4	C 4
A18	A 2	C 5	C 6
A19	A19	C 6	C 7
A20	A20	C 8	C 8
B 1	B 1	C 9	C 9
B 3	B 3	C11	C11
B 4	B 4	C13	C13
B 5	B 5	C17	C17
B 8	B 8	C18	C18
B 9	B 9	C19	C 1
B10	B20	C20	C20

C 1	C 1		
C 2	C 2		
C 3	C 4		
C 4	C 5		
	C 6	B 4	
C 8	C 9		
C10	C11		
C11	C12		
C12	C13	C 6	D 8
	C14		A 8
C14	C18		
C15	C19		
	C20	A16	
		C10	B16
		C15	C12
		C16	C 8
		D 2	A 1
		D 6	B 3
		D 9	A 6
		D10	A 8
		D11	C11
		D15	D11
		D16	A15

ANALYTICAL THINKING

GRADE 4 GRADE 6 GRADES 7,9 GRADE 11

A 1	A 6		
A 2	A 7		
A 3	A 8		
A 4	A 9		
A 5	A10		
A 6	A11		
A 7	A 1		
A 8	A 2		
A 9	A 3		
A10	A 4		
A11	A 5	B 8	B 2
		A 5	A 1
		A 6	A 3
B 1	B 8		
B 2	B 9		
B 3	B 7		
B 4	B 1		
B 5	B 2	A 1	
B 6	B 3		
B 7	B 4		
B 8	B 5	A 2	
B 9	B 6	A 3	
B10	B10		
B11	B11		
		B 1	B 5
		B 2	B 6
		B 3	B 7
		B 4	A 6
		B 5	A 7
		B 6	C 1
		B 7	B 1
C 1	C 1		
C 2	C 2		
C 3	C 3		
C 4	C 4		
C 5	C 5		
C 6	C10		
C 7	C11		
C 8	C 6		
C 9	C 7		
C10	C 8		
C11	C 9		
		C 3	C 4
		C 4	C 5
		C 5	D 8
		C 7	C 3
		D 1	D 3
		D 2	D 4
		D 3	D 5
		D 4	A 4
		D 5	A 5
		D 8	D 7

COMPUTER LITERACY

(Census tested; all forms identical)

GRADE 4 GRADE 6 GRADES 7,9 GRADE 11

1	2	2	2
2	3	1	1
		3	3
3	4		
4	5	4	4
5	6		
6	7		
7	8		
8	9		

There are seven noncognitive items which are the same at all grade levels.

ARTS AND HUMANITIES

GRADE 4 GRADE 6 GRADES 7,9 GRADE 11

		A 2	B 1
		A 4	C 4
		A 5	D 4
		A 7	A 8
		A 8	D 7
		A12	D12
	A 2	B 2	
	A 4	D 4	A 3
	A 5	D 5	A 4
	A 6	D 6	A 5
	A 7	D 7	A 6
A 4	A 7		
	A 9	B 8	
	A10	B11	
	A11	A10	
B 1	B 2		D 1
B 2	B 3		
	B 4	D 1	
	B 5	B 6	
B 5	B 6		
B 6	B 7		
	B 8	B 7	
	B 9	A11	
B10	B11		
	B12	C12	B12
		B 5	D 2
		B 9	B 5
		B10	C10
		B 1	
C 2	C 2		
	C 3	D 3	
	C 4	A 3	
C 4	C 5	B 4	
	C 6	C 4	
C 5	C 7		
C 6	C 8		
	C 9	A 6	A 7
C 8	C10		
C10	C12		
		C 1	A 2
		C 2	C 1
		C 3	B 2
		C 5	D 3
		C 7	C 7
		C 8	D 6
		C 9	C 8
		C10	A11
		C11	B10
		D 2	C 3
		D 9	D 5
		D11	C11

UNDERSTANDING OTHERS

Items are similar at Grades 4 and 6 with the addition of one situation at Grade 4. Items at Grades 7, 9 and 11 are different from each other and from Grades 4 and 6.

CITIZENSHIP/SOCIAL STUDIES

GRADE 4	GRADE 6	GRADES 7,9	GRADE 11
		A 2	B10
		A 7	A 2
	A 5	A 8	
A 2	A 8		
		A 9	D 4
		A14	D12
		A15	C15
		B14	D14
B 1	A14		
	A15		
	A16	A 1	
	B 1	B 8	D 5
	B10	B 1	
B 5	B13	C 4	
B 6	B16	A 4	
		B 3	C 9
		B 4	A 3
		B10	B 2
		B13	D11
		B15	C13
	C 1	B 9	
	C 2	B 6	
C 1	C 4		
C 2	C 5	A 6	
C 3	C 7		
C 4	C 8		
	C 9	C15	B14
C 6	C10		
	C15	C 8	A 6
C 7	C16		
		C10	D15
		C14	B13
		D 6	B 6
		D13	A13
		D15	B15

SCIENCE AND TECHNOLOGY

GRADE 4	GRADE 6	GRADES 7,9	GRADE 11
A 1	B 7		
A 2	A 1		
A 3	A 4		
A 4	A 5		
A 5	A 6		
A 7	A 9		
A 8	A10		
A10	A12		
B 1	A 2		
B 2	A 3		
B 3	A 7		
		A 1	C11
		A 7	A 7
		A 8	A 2
		A 9	D 4
B 4	B 1		
B 5	B 3	B 8	
B 7	B 5		
B 8	B 6		
B 9	B10		
B10	B11	A 4	D 2
		B 4	D11
		B 6	D 5
C 1	C 1		
C 2	C11		
C 3	C 3		
C 4	C 4	D10	
C 5	C 5		
C 6	C 6		
C 7	C 7		
C 8	C 8		
C 9	C 9		
C10	C12		
		C 3	B 6
		C 8	A 6
		C10	A10
		D 7	A 5

CITIZENSHIP/SOCIETAL RESPONSIBILITY
(Census tested; all forms identical)
(No items identical at Grades 7, 9 and 11)

GRADE 4 GRADE 6

2	2
3	3
4	4
5	5
7	7

SCIENCE AND TECHNOLOGY ATTITUDES
(Census tested; all forms identical)

GRADE 4 GRADE 6 GRADES 7,9 GRADE 11

1	1	1	1
2	2	2	3
		3	4
		4	5
	3(r)	5	6
4	4	6	8
5	5		
6	6		

HEALTH KNOWLEDGE

GRADE 4 GRADE 6

A 1	A 5
A 3	A 8
A 4	A 9
A 8	A11
A 9	A13
A11	A15
B 1	B 2
B 3	B 5
B 4	B 6
B 9	B12
B10	B14
B11	B17
C 2	C 2
C 4	C 4
C 7	C 9
C 8	C11
C 9	C13
C11	C17

ENVIRONMENT KNOWLEDGE

GRADE 4	GRADE 6	GRADES 7,9	GRADE 11
A 1	A 1		
A 2	A 2		
	A 3	A 1	
A 3	A 4	A 3	
A 4	A 5		
	A 6	A 6	
A 6	A 7		
A 7	A 8		
A 8	A 9		
		A 8	D 8
B 1	B 1		
	B 2	C 7	
B 3	B 3	A 7	
B 6	B 6		
B 7	B 7		
B 8	B 8	A 9	C 1
B 9	B 9	D 7	
		B 2	D 1
		B 3	B 6
		B 5	A 8
		B 9	D 3
C 1	C 1	B 4	
C 2	C 2		
C 3	C 3		
C 4	C 4		
C 5	C 5		
C 6	C 6	B10	
C 7	C 7		
C 8	C 8	C 4	
C 9	C 9	D 4	
		C 5	B10
		C 6	D 4
		C 9	A 7
		D 2	D 5
		D 3	C10
		D 6	A 9
		D 8	C 5

ENVIRONMENT ATTITUDES

(Census tested; all forms identical)

GRADE 4	GRADE 6	GRADES 7,9	GRADE 11
1	1		
2	2	1	
3	3	2	
		3	3
4	4		
5	5	5	
6	6		

HEALTH AND SAFETY PRACTICES

GRADE 4	GRADE 6	GRADES 7,9	GRADE 11
A 1	A 1	B 1	
A 2	A 2	B 4	C 2
A 3	A 3	D 9	D11
A 4	A 4	A 5	D 4
	A 5	A 6	
A 6	A 6	B 7	D 8
A 7	A 7		
	A 8	A 9	C10
	A 9	B10	
A10	A10	A11	B12
A11	A11	A12	
B 1	B 1	B 2	A 2
B 2	B 2	A 3	A 4
B 3	B 3		
	B 4	D 4	C 6
B 4	B 5		
B 5	B 6	B 6	B 3
B 6	B 7		
B 7	B 8	C 7	B14
B 8	B 9	D11	C13
B 9	B10	B11	B13
B10	B11(r)	B 9	B10
B11	B12	A 4	A 5
	C 1	B 3	A 3
	C 2	C 1	
C 3	C 3	C 5	A 7
	C 4	C 6	A 8
C 5	C 5		
C 6	C 6		
	C 7	B12	
C 7	C 8	D 7	C 8
C 8	C 9	C 9	C 9
C 9	C10		
	C11	D12	B11
C11	C12	C12	
		A 1	A 1
		A 7	C 7
		A10	C11
		B 5	B 2
		B 8	B 6
		C 2	B 5
		C 3	B 1
		C 4	C 3
		C 8	A14
		C10	B 9
		C11	D13
		D 2	C 1
		D 3	A 6
		D 5	B 4
		D 6	D 7
		D 8	A10
		D10	B 8

TEACHER QUESTIONNAIRE

School Name _____ District Name _____

INSTRUCTIONS: Please respond to the following items by blackening the circles with a No. 2 PENCIL ONLY. The information you provide will be an important part of the school assessment. Information received will be aggregated and reported in terms of relationships to student achievement. No individual information will be reported.

- | | |
|---|---|
| <p>1. Are you a classroom teacher; that is, do you spend more than 50% of your school time in classroom instruction?
 ☐ A Yes ☐ B No</p> <p>2. In your teaching situation, how satisfied are you with your relationships with parents and parent groups?
 ☐ A Very Satisfied ☐ C Somewhat Dissatisfied
 ☐ B Somewhat Satisfied ☐ D Very Dissatisfied</p> <p>3. Which of the following best describes your level of formal education?
 ☐ A No degree ☐ D Master's degree plus one year
 ☐ B Bachelor's degree ☐ E Doctor's degree
 ☐ C Master's degree or equivalency</p> | <p>4. I teach mostly:
 ☐ A Self-contained elem. ☐ F Reading
 ☐ B Business ☐ G Social studies
 ☐ C English ☐ H Science
 ☐ D Foreign language(s) ☐ I Health/Phys. ed.
 ☐ E Mathematics ☐ J Other</p> <p>Items 5 and 6 for elementary teachers only.</p> <p>5. What percentage of your students' parents attend scheduled parent-teacher conferences?
 ☐ A 0-20 percent ☐ D 61-80 percent
 ☐ B 21-40 percent ☐ E 81-100 percent
 ☐ C 41-60 percent</p> <p>6. Who is in charge of this building?
 ☐ A Principal ☐ C Head Teacher
 ☐ B Elementary Supervisor ☐ D Other</p> |
|---|---|

NOTE: IF YOU TEACH IN A BUILDING THAT DOES NOT HAVE A PRINCIPAL, respond to the "Principal" items that follow in this questionnaire in terms of the person who serves in lieu of the principal, such as a head teacher, supervisor, elementary director or secondary director.

For items 7 to 26, blacken the circle that best tells how you feel about each statement.

SA=Strongly Agree
 MA=Mostly Agree
 NN=Neither Agree nor Disagree
 MD=Mostly Disagree
 SD=Strongly Disagree

- | | |
|---|--|
| 7. Teachers and students respect the principal | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 8. The principal is knowledgeable about teaching techniques | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 9. The principal conveys to the community a positive view of the school and its program, staff and students | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 10. Teachers in this school hold consistently high expectations for all students | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 11. Teachers in this school seek better ways of teaching and learning | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 12. Teachers in this school are proud to be teachers | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 13. Teachers in this school are knowledgeable about their subject areas | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 14. Students in this school are interested in learning | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 15. Student attendance needs improvement | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 16. Teacher attendance needs improvement | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 17. People in this school solve problems rather than just talk about them | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 18. Announcements given through the public address system are disruptive to my classroom teaching | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 19. There are student-initiated disruptions of my classes | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 20. Students are held responsible for the consequences of their behavior | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 21. Too much time is spent on discipline | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 22. Too little support is provided by administrators on discipline | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 23. Assessment information is used by the school to set improvement priorities | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 24. Teachers know the school goals established for this year | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 25. Teachers have open channels of communication with district administrators | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |
| 26. The physical facilities of this school are generally unpleasant and poorly maintained | ☐ SA ☐ MA ☐ NN ☐ MD ☐ SD |

For items 27 to 50, blacken the circle that best tells how you feel about each statement

A=Always
B=Almost Always
C=Frequently
D=Occasionally
E=Rarely or never

27. A positive feeling permeates the school. (A B C D E)
28. The principal runs effective meetings; that is, he/she has a clear agenda, limits discussion to relevant topics and adheres to the time frame. (A B C D E)
29. The principal encourages me to solve my own work problems but is available to advise me if needed. (A B C D E)
30. The teachers feel this school is run in an orderly fashion without being overly restrictive. (A B C D E)
31. The principal talks with us frankly and openly. (A B C D E)
32. Teachers handle general student discipline in a reasonable way. (A B C D E)
33. Classroom atmosphere in this school is conducive to learning. (A B C D E)
34. Teachers praise students for good performance. (A B C D E)
35. Teachers are cooperative and supportive of each other. (A B C D E)
36. Teachers treat students with respect in this school. (A B C D E)
37. My instruction is disrupted by students being excused from class for various activities. (A B C D E)
38. I have to spend too much time on noninstructional duties. (A B C D E)
39. Teachers, students and parents are given opportunities to provide input into the curriculum development process. (A B C D E)
40. Parents provide support to teachers on discipline. (A B C D E)
41. Students treat teachers with respect. (A B C D E)
42. Students are well-behaved in this school. (A B C D E)
43. Students obey the school's rules. (A B C D E)
44. Students complete assigned homework before coming to class. (A B C D E)
45. Released time and financial support for in-service training are provided for teachers. (A B C D E)
46. During parent-teacher conferences there is a focus on factors directly related to student achievement and basic skill mastery. (A B C D E)
47. Adequate materials and supplies necessary for instruction are available to me. (A B C D E)
48. This school is a safe and secure place to work. (A B C D E)
49. I look forward to coming to work. (A B C D E)
50. I attend conferences related to the areas I teach. (A B C D E)

51. What is your average class size? (Exclude supervisory duties such as study hall.)

0	0
1	1
2	2
3	3
4	4
5	5
6	
7	
8	
9	

52. Which grade level(s) do you presently teach?

(Mark all that apply)

- K ☐
- 1 ☐
- 2 ☐
- 3 ☐
- 4 ☐
- 5 ☐
- 6 ☐
- 7 ☐
- 8 ☐
- 9 ☐
- 10 ☐
- 11 ☐
- 12 ☐

53. Within an average week, how many different students do you teach?

0	0	0
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	
7	7	
8	8	
9	9	

54. How many formal classroom observations of your instruction are made each year?

- (A) None (D) Three
- (B) One (E) Four or more
- (C) Two

55. How many buildings does your principal supervise?

- (A) One (C) Three
- (B) Two (D) Four or more

56. Which of the following best describes your current teaching assignment?

- (A) I teach one group of students for the entire day.
- (B) I teach one subject to two or more different classes of students.
- (C) I teach two or more subjects to two or more groups of students.